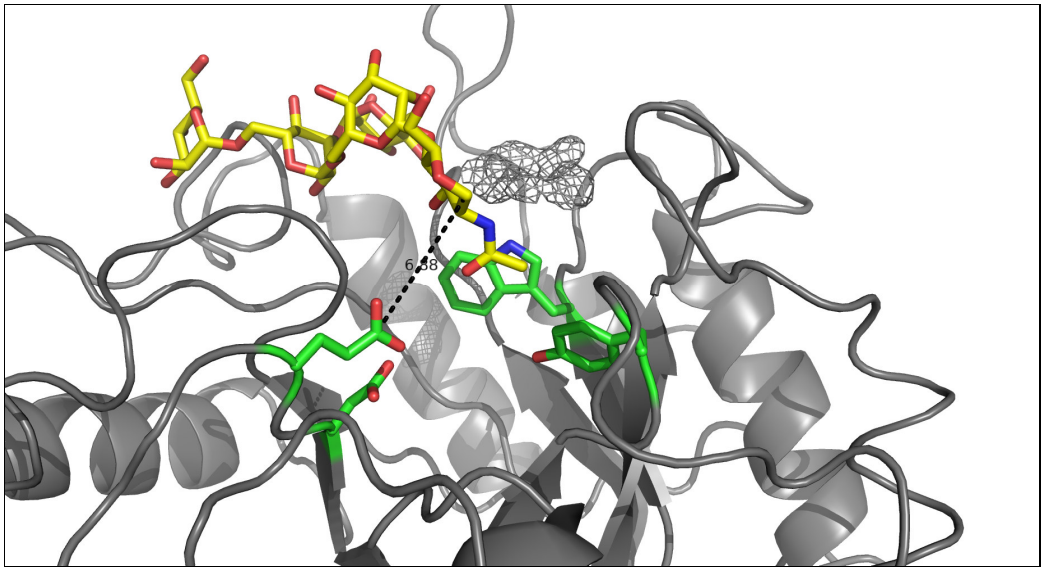


Mutational Engineering, Growth Selection, Expression, Purification and Crystallization of Ligand Bound Endo-β-*N*-acetylglucosaminidase T (Endo-T) from *Hypocrea jecorina*

Andreas Digre

Department of Molecular Biology, Swedish University of Agricultural Sciences, Uppsala, Sweden

Independent Project in Biology D, 30 HEC, EX0564

Spring 2010, SLU

Supervisors: **Mats Sandgren**, Ingeborg Stals, Saeid Karkehabadi, Jerry Ståhlberg, Nils Mikkelsen and Henrik Hansson

---

Keywords:

Endo-T ♦ Endo-β-*N*-acetylglucosaminidase ♦ Deglycosylation ♦ EC 3.2.1.96 ♦ Crystallography ♦ Ligand bound structure ♦ High mannose N-glycan ♦ HPAEC-PAD

Submitted: June, 2010

Thesis contents

Abstract

Recently, one of the first fungal-expressed, deglycosylating, Endo-β-*N*-acetylglucosaminidases was found in the extracellular medium of soft-rot ascomycete *Hypocrea jecorina* (a.k.a. *Trichoderma reesei)* belonging to glycoside hydrolase family 18 (GH18). It was named Endo-T and has been shown to possess similar substrate specificities as Endo-H from *Streptomyces plicatus*, a deglycosylating enzyme, frequently used in the field of glycoproteomics. In this study an oligomannosidic N-glycan was introduced in the active site of crystallized Endo-T under acidic pH conditions below pH 3. The ligand-containing crystal diffracted on a synchrotron x-ray source to a resolution of 1.65Å. The resulting Endo-T structure was found to contain a bound ligand consisting of a hydrolyzed Manα1-6(Manα1-3)Manα1-6Manβ1-4GlcNAc N-glycan, lacking the aspargine linked N-acetylglucosamine residue of N-glycans. Furthermore, electron density was missing for several of the distal glycone mannose residues. The anomeric carbon of the distal N-acetylglucosamine residue was found to be positioned 6.88Å from the proposed catalytic amino acid residue, Glu131, indicating a descending motion during hydrolysis. An unidentified electron density was found after the last structure refinement, apparently shaped as a furanose ring, next to the N-glycan ligand in the active site around unit 8 of the $(\beta/\alpha)_8$-barrel. Proximal substrate positioning and the loop-structure of Endo-T suggests aglycone docking centered over $(\beta/\alpha)_8$-barrel unit 5.

Introduction

Glycosylation

Glycosylation is the covalent attachment of carbohydrates to non-carbohydrate organic compounds, e.g. proteins and lipids, catalyzed by glycosyltransferases[1, 2]. The biological functions of the oligosaccharides attached are many and diverse spanning from sterical protection from degrading enzymes and structure stabilization to hormonal, cell-cell and cell-matrix interactions[3]. The reason for the diverse range of biological functions of the carbohydrates is a wide variety of structural combinations. There are, for instance, 10^{12} different isomers of a hexasaccharide consisting of only six monosaccharide residues[4].

N-linked glycoproteins

Glycoproteins are a common type of glycoconjugates where the glycan is covalently attached to the polypeptide backbone of a protein in an O-linked or N-linked matter. The O-linked glycans are attached to the hydroxyl group of the amino acids serine and/or threonine while the N-linked glycans are bound to the amide group of the amino acid aspargine[5]. N-glycans are most often involved in interacting with the environment as a recognition signal and all of these glycans contain a specific pentasaccharide core: Manα1-6(Manα1-3)Manβ1-4GlcNAcβ1-4GlcNAc, where the two N-acetylglucosamines are located at the reducing terminal attached to the aspargine[5]. The differences starts with the glycan residues linked further onto the pentasaccharide core resulting in three different types of N-glycans: Complex, high mannose and hybrid N-glycan chains[5]. The complex types of N-glycans contain no additional mannose residues but the three mannose residues included in the pentasaccharide core and instead contain other glycan residues linked to the three mannose and the proximal N-acetylglucosamine residues of the core[5]. The high mannose type N-glycans have only mannose residues attached, purely to the two distal mannose residues of the pentasaccharide core. The hybrid group contains both characteristics of complex and high mannose type N-glycans[5].

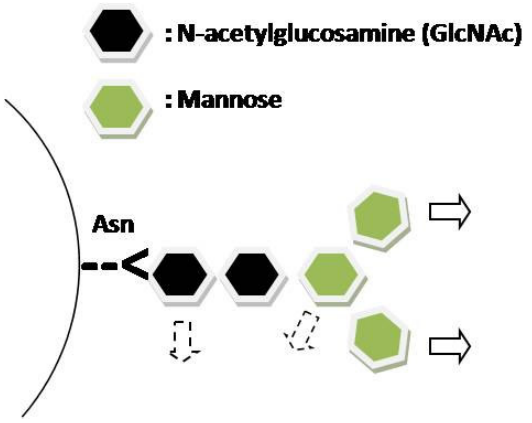


Figure 1. **Pentasaccharide core of N-linked glycans.** The pentasaccharide core consisting of a Manα1-6(Manα1-3)Manβ1-4GlcNAcβ1-4GlcNAc molecule attached to an aspargine residue on the aglycone protein. Arrows indicate the possible extensions of glycan residues (dotted arrows indicate single monosaccharide extensions only found in complex and hybrid N-glycans)[5].

Deglycosylation

Deglycosylation is the process of removing sugar residues from glycosylated organic compounds by glycosidic bond hydrolysis. The glycosidic bond, which links the carbohydrate residues to other carbohydrates or other non-carbohydrate organic compounds, is cleaved in a hydrolytic reaction comprising two essential catalytic residues: a proton donor and a nucleophile/base, most often aspartic acid and/or glutamic acid amino acid residues in the active site of the deglycosylating enzyme[6, 7]. Different positions of the base residue with regard to the carbohydrate anomeric carbon facilitate either retention or inversion of the anomeric configuration after hydrolysis. The retaining mechanism is characterized by a short distance between the base residue and the anomeric carbon while the inverting mechanism often demonstrates a larger distance between the two (~4.5 Å longer) to provide room for the attacking water molecule creating the inverted anomeric configuration[8].

Glycoside hydrolases – classification systems

Glycoside hydrolases (EC 3.2.1.x) are the enzymes responsible for degrading the complex and diverse carbohydrate structures on glycosylated proteins. As carbohydrates are both structurally complex and involved in a multitude of biological functions there is a huge number of various glycoside hydrolases expressed by all realms of life. The IUB (International Union of Biochemistry) Enzyme Nomenclature (1984) was the first established classification model for enzymes based on the biochemical reaction performed by the enzyme and on its specificity towards substrates. This classification system is a good way of avoiding redundant inventions of names for the same enzyme and creating a universal system. A problem with the system is that it does not take into account the structural properties of the enzymes as the reaction and substrate-specificity often does not reflect the three dimensional structure of the protein which is a very common phenomena amongst the glycoside hydrolases. Another problem is that the classification is based on a "one kind of enzyme, one kind of substrate" way of thinking even though many enzymes demonstrate a wide specificity towards several substrates[9]. As it had previously been shown that structural properties were directly related to the amino acid sequence of proteins[10] a new classification system was created which is based on peptide sequence similarities of carbohydrate-processing enzymes[9, 11-15]. This system does not only expose structural and evolutionary relationships between the different enzymes but does also provide information about the reaction mechanisms as the IUB system[9]. The carbohydrate-processing enzymes have been classified into numerous sequence-related families, where the glycoside hydrolases currently include 118 families (CAZy)[16], and these families are frequently updated as sequences and three-

dimensional structures of additional enzymes are being solved continuously. A database have been established which lists these families of carbohydrate-processing enzymes, called CAZy (Carbohydrate Active Enzymes, http://www.cazy.org/)[16].

Endo-T

Endo-β-N-acetylglucosaminidases

Endo-β-*N*-acetylglucosaminidases (IUB: EC 3.2.1.96, ENGases, Endo:s) are a group of endoglycoside hydrolases which catalyze the hydrolysis of the glycosidic bond between the two proximal N-acetylglucosamines linked to the aspargine in N-linked glycans leaving one N-acetylglucosamine residue with the aglycone (aspargine part) and the other becomes the reducing end of the detached N-glycan (glycone)[21, 45]. A two-step retaining substrate assisted mechanism is believed to be utilized (Figure 2). In the first step, the acetamido group of the non-asparagine bound N-acetylglucosamine residue of the substrate makes a nucleophilic attack on the anomeric carbon and forms an oxazoline intermediate. In the second step, the attack of a water molecule on the anomeric carbon of the substrate completes the cycle of hydrolysis[20, 52-56].

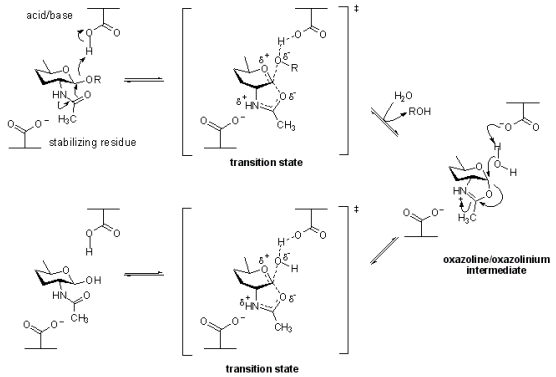


Figure 2. **Two-step substrate assisted hydrolysis** (www.CAZypedia.org). The formation of an oxazoline intermediate by the non-aspargine linked N-acetylglucosamine residue of the substrate, assisting in a two-step hydrolysis of an N-glycan substrate together with a proton donor residue and a stabilizing residue[20, 52-56].

A carboxylate of an amino acid residue is believed to stabilize the oxazoline intermediate during the transition states of the reaction and the catalytic proton donation is believed to be performed by a glutamic acid residue[16]. Members of this group of enzymes are found in family 18 (GH18) and 85 (GH85) of glycoside hydrolases[16]. Four members of Endo-β-*N*-acetylglucosaminidases are present in family GH85 expressed by both eukaryotes and prokaryotes: Endo-M from *Mucor hiemalis*[17], Endo-CE from *C. elegans*[18], Endo-A from *Arthrobacter protophormiae*[19] and Endo-D from *Streptococcus pneumoniae*[20]. GH18 are related to chitin-degrading chitinases and contained until recently Endo-β-*N*-acetylglucosaminidases solely expressed by prokaryotes including Endo-H from *Streptomyces plicatus*[21], Endo-F1, F2, F3 from *Elizabethkingia meningoseptica*[22] and Endo-S from *Streptococcus pyogenes*[23]. Recently, however, new members of GH family 18 have been discovered to be expressed by both ascomycete and basidiomycete fungi comprising Endo-T from *Hypocrea jecorina* and its homologous proteins[24] and Endo-FV from *Flammulina velutipes*[25] respectively.

An endo-β-N-acetylglucosaminidase found in H. jecorina

The Endo-β-*N*-acetylglucosaminidase that recently was found to be expressed and secreted by the mesophilic soft-rot ascomycete *Hypocrea jecorina*, formerly known as *Trichoderma reesei* (hence the "T" in Endo-T), by Stals and her colleagues has been shown to enzymaticly process N-linked oligomannosidic, phosphorylated and hypermannosylated glycoproteins[24]. Endo-T has low sequence homology to other protein members of GH18 family but has similar substrate-specificities as two other GH family 18 group members, Endo-H and Endo-F1[21, 22, 24]. Endo-T consists of 287 amino acids (after N- and C-terminal processing; Ingeborg Stals, Ghent University, Belgium,

personal communication). It has a pI of 4.33 (ProtParam, ExPASy Proteomics Server) and a mass of approximately 30 kDa (according to SDS-PAGE analysis but with a theoretical mass of 31.755kDa) depending on the grade of N- and C-terminal processing (Ingeborg Stals, Ghent University, Belgium, personal communication). The homologue Endo-H from *Streptomyces plicatus,* found 1974 by Tarentino and his colleagues[21] is nowadays frequently used as a deglycosylation reagent in glycoprotein research to study the structure and functions of N-linked carbohydrates[26, 27]. Another area of usage is the removal of oligosaccharides from glycosylated proteins to enable biochemical characterization and crystallography studies of the aglycone protein (personal communication). As Endo-H is fairly expensive to purchase at the moment as it is patented and sold extensively overpriced by Roche Diagnostics, Endo-T has emerged as a good candidate substitute since it has similar substrate specificities as Endo-H and can be produced at a high rate due to the high enzyme-producing properties of *Hypocrea jecorina*. Hence, Endo-T has the potential of significantly reducing the cost for this type of enzymatic deglycosylation treatment worldwide.

Structure biology of Endo-T

The structure of Endo-T expressed by *Hypocrea jecorina* was solved by Dr. Saeid Karkehabadi and Dr. Mats Sandgren at the department of molecular biology, Swedish University of Agricultural Sciences, Uppsala, Sweden (personal communication) with help from a former bachelor project student Mia C. Hertzberg, to a resolution of 1.3 Å by multiple anomalous dispersion (MAD) methods, using bound zinc molecules for the phasing[28](Figure 3). The fold of the protein is the common TIM $(\beta/\alpha)_8$-barrel shared by the ENGases but with some unique folding differences to the Endo-β-*N*-acetylglucosaminidase and chitinase

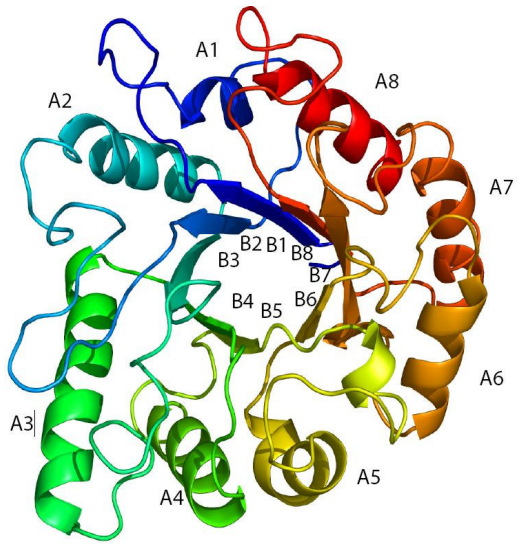


Figure 3. **Cartoon representation of the three-dimensional structure of Endo-β-N-acetylglucosaminidase T from Hypocrea jecorina**. The $(\beta/\alpha)_8$-barrel structure is colored from blue to red starting at the N-terminous of the enzyme. The eight α-helix (A) and β-barrel (B) units of the $(\beta/\alpha)_8$-barrel are labeled in the figure.

structures of GH family 18 (Saeid Karkehabadi et al., personal communication). The $(\beta/\alpha)_8$-barrel consist of eight β-strand/loop/α-helix units (Figure 3) where the catalytic site is positioned on top of the core of the barrel-structure (Saeid Karkehabadi et al., personal communication). Aspartic acid 129 and glutamic acid 131 of Endo-T are found to be conserved amino acids within GH family 18, and these amino acids are thought to be catalytic residues as the mutational substitution of these amino acids in Endo-H resulted in loss of hydrolytic activity[29]. Tyrosine 195 and tryptophan 259 are also believed to have important roles where Tyr 195 has been shown to be important in homologue enzymes, aiding the substrate binding[30] and Trp259 is believed to stabilize the enzyme-substrate interaction[28]. Endo-T has been shown to be glycosylated with a single N-acetylglucosamine residue on aspargine 70 and 240 [28].

X-ray crystallography

Crystallization

Crystals are solidified ordered aggregates of molecules and are formed under supersaturation, i.e. a non-equilibrium state[31]. At a non-equilibrium state, a solution contains more of a dissolved molecule than could be dissolved under normal conditions and therefore results in solidification of that molecule until the equilibrium concentration of saturation is reached[31]. The formation of crystals starts with nucleation, when the molecules diverge from the single molecule solution state into ordered multimolecular aggregates[31]. Nucleation is followed by crystal growth, where the molecules are obliged to join the multimolecular nucleations instead of roaming free in the solution[31]. Crystallization of proteins is a crucial step in solving their three-dimensional structure and is utilized extensively in the fields of biology. A problem with protein crystallization is that macromolecules can under supersaturation, except from forming crystals, also form disordered aggregates which are useless[31]. The art of generating protein crystals involves many different parameters including protein sample purity, pH, temperature, precipitants and buffers. The purity of the protein sample used for crystallization experiments is critical as protein crystals are aggregates with perfectly ordered identical protein units and are extremely sensitive to foreign proteins or structurally deviating protein variants[31]. Precipitants are compounds used in the crystal solution to change the interactions between the protein and the surrounding solvent to promote protein precipitation out of the solution into crystals[31]. There are different types of precipitants including salts, organic solvents and polymers, where the salts ammonium sulphate and sodium chloride and the polymer polyethylene glycol (PEG) are the most commonly used. The methods used for protein crystallization includes the most frequently used vapor diffusion method and other techniques such as dialysis, the batch method and free interface diffusion[31]. In the vapor diffusion method a droplet of typically 1-10µl highly concentrated protein sample is placed on a glass lid or pedestal and mixed with equal amounts of crystallization solution containing buffer and precipitants[31]. The drop is then placed in a sealed environment containing a well filled with 500-1000µl crystallization solution. As the concentration of the crystallization solution components is lower in the droplet compared to the well solution, vapor diffusion brings the environment to equilibrium by supersaturation of the droplet and hopefully crystal formation[31].

X-ray diffraction

Structure determination from protein crystals by X-ray diffraction is possible because of the ability of electrons in the crystal to scatter X-rays. When an X-ray beam hits the electrons of a protein crystal it becomes scattered and creates a specific diffraction pattern, the combinatorial diffraction from probably millions of electron induced diffractions[32]. This pattern of scattered X-ray can be reversibly calculated back to the positions of its scattering sources, the electrons of the protein crystal and by retracing the positions of the protein electrons the spatial occupancy of all atoms in the protein can be determined, i.e. the protein electron density map[32]. The crystal construct is also vital for protein structure determination. The diffraction signals from the electrons of a single protein molecule are undetectable[32]. The ordered crystal lattice organization of repeated protein molecules creates an amplification of the diffraction signals and enables diffraction pattern detection[32]. However, first of all the protein of interest must be crystallized into good quality protein crystals. The protein crystals must then be flash frozen in liquid nitrogen (100 K) to prevent X-ray-induced radiation damage of the

protein crystal in the x-ray beam. The flash freezing itself can cause crystal-damage by crystalline ice formation[31]. This is however prevented by coating the crystals with cryoprotectant solution which has a lower concentration of ice-forming water molecules[31]. The protein crystal is subsequently transferred onto a goniometer on an X-ray machine which can accurately position and rotate the crystal in relation to the X-ray beam. The protein crystal is then irradiated with an X-ray beam originating from a synchrotron or home lab generator X-ray source. The pattern of the diffracted X-ray beam was formerly detected by an X-ray film but is nowadays detected by modern image plates and area detectors[32]. In order to collect a complete diffraction data set, containing information about the density of almost all electrons of the protein molecule, the crystal has to diffract X-ray in multiple rotation angles[32]. The three-dimensional electron density map of the crystal unit cell is then calculated from the intensity and positioning of the diffraction pattern spots of the crystal in different orientations[32]. The important information needed to determine the protein structure is obtained from the intensity of the diffraction spots which are related to the amplitude of the electromagnetic wave of the diffracted X-ray[32]. To be able to calculate the electron density map both amplitudes and phase angles of the electromagnetic waves must be known[32]. While the amplitude can be calculated from the spot intensity, the phase angles can not be extracted from the collected X-ray diffraction dataset[32]. This problem, known as the phase problem is solved by the use of three existing methods: 1) The isomorphous replacement method[32] where diffraction pattern intensity differences of protein-linked heavy atoms are used as known factors to determine the phase angles, 2) The anomalous scattering methods[32] which utilize the deviating X-ray scattering by electrons in the inner shells of the electron cloud of native heavy atoms of the protein at certain wavelengths for phase angle determination or 3) The molecular replacement method[32] where the phase information from the solved structure of a similar protein is combined with the diffraction spot intensities of the unsolved structure to obtain the electron density map. The molecular replacement method is mostly applied for solving structures of ligand-bound or mutated proteins where the apo- or wild type structure is already solved. After obtaining an electron density map, different computer programs are used to build the protein molecule that both fits well inside of the electron density and fits the correct geometry of a protein molecule, such as bond angles and bond lengths[32]. The protein structure model will be refined until the researcher determines the structure model to be of acceptable quality. The quality of the crystal, i.e. the concordance between the diffraction data and the structure model, can be demonstrated by the R and R_{free} factor[32].

High performance an-ion exchange chromatography with pulsed amperometric detection (HPAEC-PAD)

A high performance an-ion exchange (HPAE) method was developed due to the lack of simple and effective alternatives for carbohydrate-analyzing chromatography methods[33]. Together with its pulsed amperometric detection (PAD)[33] it can separate and quantify extremely small amounts in the order of picomole carbohydrate in a sample, while at the same time being highly tolerant against sample solution impurities. Neutral carbohydrates are actually very weak acids with pK_a values between 12 and 14 and this is exploited in HPAE chromatography[33]. A strong an-ion exchange stationary phase is used at extremely high pH to ionize the carbohydrates and enable a highly discriminating separation by charge[33]. Because of the high pH, neutral and positively charged compounds are conveniently eluted at or close to the void

volume of the column[33]. The column material of these anion exchangers consist of highly pH stable (0-14), robust, polymeric, nonporous micro-beads[33]. For reduced sugars however, a macroporous polymeric resin is used which has 45 times stronger an-ion strength to be able to bind these weakly anionic molecules effectively[33]. The pulsed amperometric detection is an extremely sensitive detection-method for carbohydrates as it can distinguish sugar-sample amounts as low as 10 picomoles from the background noise[33]. This method of detection is characterized by repeated sequences of three different oxidation potentials (P1-3)[33]. Firstly, the eluted carbohydrates are detected via oxidation of the carbohydrates (P1) at the surface of a gold electrode[33] creating an electric current that is measured by the detector. During the carbohydrate oxidation, poisonous products are formed at the gold surface[33] which is removed by raising the oxidation potential resulting in an oxidation of the gold electrode (P2). Finally the gold oxide is reduced back to gold by lowering the potential (P3) and then another oxidation-sequence can start[33]. The measured amounts of oxidized compounds (mainly sugar compounds) are collected as data files. These data files can then be used by certain computer programs for quantitative and purity analysis of the characterized carbohydrate containing sample[33].

Ligand-bound Endo-T structure

The major aim of this independent master project was to crystallize and solve the structure of *H. jecorina* Endo-β-*N*-acetylglucosaminidase-T with a ligand bound in the active site of the enzyme. A bound ligand provides vital information about enzyme-substrate interactions including interactions with the glycone part of the ligand as well as the aglycone (carbohydrate-protein) region. The glycone interactions contain mechanistic information regarding the residues in close proximity to the active site, and the binding-preferences concerning the diverse structure of the distal complex of oligosaccharide residues. The aglycone region interactions include the area around the two proximal N-acetylglucosamines linked to the aspargine and the amino acid interactions between the glycopeptide/glycoprotein and Endo-T. The exploration of these interactions will facilitate the understanding of the biological role of ENGases and enable future designs of glycoside hydrolases with desirable properties. In this study attempts were made to crystallize Endo-T with a ligand in the active site via enzyme inactivation by either extreme pH conditions or the design of inactivated Endo-T mutants. The selection of inactivated *H. jecorina* Endo-T mutants failed to give rise to healthy mutant *H. jecorina* clones with acquired Endo-T overexpression. However, Endo-T ligand complex crystals, using wild type enzyme, were produced under extreme pH-conditions i.e. pH 2.4-3.0, and these crystals were used for ligand complex structure determination. The ligand comprises a processed and incomplete Manα1-6(Manα1-3)Manα1-6Manβ1-4GlcNAc N-glycan ligand. In spite of the low pH in which the crystals were formed, inspection of the electron density map clearly showed that Endo-T was not completely inactivated since the substrate was processed. Electron density for a furanose ring was also detected in the active site of the ligand-bound Endo-T.

Materials and methods

Ligand preparation and analysis

Preparation of $(GlcNac)_2$-Man_{5-9} N-glycans from RNase B

The N-glycans were generally prepared as previously described[34]. 50 mg/ml of Ribonuclease B (RNase B, Sigma-Aldrich®) was dissolved in denaturing solution (1% SDS + 0.5M β-mercaptoethanol + 0.1M EDTA) and

incubated for 30 min at room temperature. The RNase B solution was then diluted in 178mM phosphate buffer (pH 8.6), boiled for 5 min at 100ºC and subsequently cooled down to room temperature. After addition of 0.75% triton X-100, the solution was incubated with 10 units of PNGase F overnight at 37ºC. On the following day the protein solution was spun down (15 min at 13000xg) and the remaining protein in the supernatant was precipitated with pre-cooled 4.3M ethanol for 1 hour at 4ºC. After another centrifugation of protein (15 min at 13000xg) the sugar of the supernatant was precipitated with pre-cooled 2.7M acetone for 10 min at -20ºC. The sugar was then spun down (15 min at 13000xg) and the pellet was resuspended in 1 ml water. The N-glycan sample was subsequently further purified with a carbograph column (Alltech™ Extract Clean™ SPE Carbo 150mg/4ml) which was regenerated with 2 ml 15.3M CH_3CN+0.135M trifluoroacetic acid and rinsed with 3 column volumes water before the loading of the N-glycan sample. The column was then washed with 5 ml water whereafter the bound sugars were eluted with 2 ml 4.8M CH_3CN. The fractions of the loading, washing and elution step were collected separately and analyzed with high performance an-ion exchange chromatography with pulsed amperometric detection (HPAEC-PAD; ICS-3000, Dionex). The elution-fraction sample was dried and resuspended in 100 µl water to create a highly concentrated ligand-sample.

Preparation of Asn-$(GlcNac)_2$-Man_{5-9} glycoaspargine from RNase B

16.7 mg/ml of RNase B (Sigma-Aldrich®) was dissolved in 200mM phosphate buffer (pH 7.0) and incubated with 11.3 mg/ml pronase (protease type XIV from *S. griseus*, Sigma-Aldrich®) overnight or for 4 days (for increased hydrolysis) at 37ºC. After the proteins of the solution were pelleted (20 min at 5000xg), the glycoaspargines of the supernatant were precipitated with pre-cooled 2.7M acetone for 10 min at -20ºC. The precipitate was subsequently spun down (20 min at 5000xg) and the pellet was resuspended in 1 ml water. The glycoaspargine sample was then further purified with a carbograph column (Alltech™ Extract Clean™ SPE Carbo 150mg/4ml), analyzed with HPAEC-PAD (ICS-3000, Dionex) and concentrated as described above ("*Preparation of $(GlcNac)_2$-Man_{5-9} N-glycans from RNase B*").

Transformation experiment

Transformation of the mutant constructs into *Hypocrea* jecorina was performed as previously described by Penttilä et. al.[35]. The *H.* jecorina Endo-T knockout strain RLP37 (Genencor) was transformed with Trichoderma expression vector pTrex3g constructs containing the *amdS* selection marker and the mutant sequence of the Endo-T gene variants D129A, Q193A, E131A and Y195A.

Clone selection process

Grown protoplast *H. jecorina* Endo-T knockout clones (RLP37 strain, Genencor) transformed with different amounts of plasmid DNA coding for the four different catalytically mutated variants of Endo-T (D129A, Q193A, E131A and Y195A) were picked with toothpicks, plated on sterile 2% glucose amdS selection plates (111mM glucose + 110mM KH_2PO_4 + 10mM acetamide + 10mM CsCl + 2.43mM $MgSO_4*7H_2O$ + 4.08mM $CaCl_2*2H_2O$ + 2.28mM citric acid + 1.80mM $FeSO_4*7H_2O$ + 139µM $ZnSO_4*7H_2O$ + 32.0µM $CuSO_4*5H_2O$ + 20.7µM $MnSO_4*H_2O$ + 32.3µM H_3BO_3 + 20g/L agar) and grown at 28ºC. Healthy growing clones from these plates were re-plated as above for a second round of amdS-based selection. The healthy growing clones from the second amdS-base selection round were grown in sterile 5 ml cultures with liquid 1.6% lactose defined media (46.7mM lactose + 2.78mM $(NH_4)_2SO_4$ + 100mM KH_2PO_4 + 4.06mM

$MgSO_4$*$7H_2O$ + 6.85mM $CaCl_2$*$6H_2O$ + 9g/L casamino acids + 2.28mM citric acid + 1.80mM $FeSO_4$*$7H_2O$ + 139µM $ZnSO_4$*$7H_2O$ + 32.0µM $CuSO_4$*$5H_2O$ + 20.7µM $MnSO_4$*H_2O + 32.3µM H_3BO_3) which were going to be screened for the expressed amounts of Endo-T with SDS-PAGE analysis (GE Phast 8-25% gradient gel, GE Healthcare).

H. jecorina growth cultures

A healthy growing clone from three of the four Endo-T mutant *H. jecorina* variants (D129A, Q193A and Y195A) and the Endo-T knock out *H. jecorina* were cultured in sterile 50 ml 2% glucose minimal media pre-cultures (111mM glucose + 27.8mM $(NH_4)_2SO_4$ + 5.41mM $CaCl_2$ + 4.98mM $MgSO_4$ + 110mM KH_2PO_4 + 9.93µM $MnSO_4$ + 18.0µM $FeSO_4$*$7H_2O$ + 15.4µM $CoCl_2$ + 9.29µM $ZnSO_4$) and subsequently in sterile 300 ml 2% lactose minimal media induction cultures (58.4mM lactose + 27.8mM $(NH_4)_2SO_4$ + 5.41mM $CaCl_2$ + 4.98mM $MgSO_4$ + 110mM KH_2PO_4 + 9.93µM $MnSO_4$ + 18.0µM $FeSO_4$*$7H_2O$ + 15.4µM $CoCl_2$ + 9.29µM $ZnSO_4$), both with shaking at 28ºC. Culture medium samples from each culture were checked for presence of Endo-T with SDS-PAGE analysis using 12% manually produced gels (Separating gel: 12% acrylamide:bisacrylamide (AA:bisAA)(37.5:1) + 375mM Tris-HCl pH 8.8 + 21.9mM ammonium persulphate (APS) + 17.2mM Tetramethylethylenediamine (TEMED); Stacking gel: 4% AA:bisAA(37.5:1) + 625mM Tris-HCl pH 6.8 + 43.8mM APS + 34.4mM TEMED). The samples and a low molecular weight marker (GE Healthcare) were mixed with sodium dodecyl sulfate (SDS)/dye sample buffer and heated 5 min at 100ºC before being loaded on the gel. The gel was then run in the electrophoresis device (Bio-Rad) for about 50 min at 200V and subsequently developed with Coomassie brilliant blue R-250 (Bio-Rad).

Endo-T purification

Avicel adsorption

300 ml extracellular medium from each of the four induction cultures, containing three Endo-T mutant *H. jecorina* clones (D129A, Q193A and Y195A) and an Endo-T knock out *H. jecorina* strain, were harvested and the fungal mycelia were removed by büchner vacuum filtration. After concentrating the filtrates by ultrafiltration at 2 bar pressure in stirred cells (Amicon, Millipore) using 10 kDa NWCO polyethersulfone membranes (Biomax-10, Millipore) equal amounts of 100mM sodium acetate (pH 5.0) and 20g of Avicel were added followed by one hour incubation with shaking at room temperature. The avicel-cellulase aggregates were subsequently spun down (30 min at 3000xg) and the supernatants with the unbound proteins were concentrated with the previously mentioned ultrafiltration method ("*Avicel adsorption*").

SDS-PAGE analysis of Avicel purified culture filtrate

SDS-PAGE analysis was then performed with 15% gels to check for the presence of bands corresponding to the size of Endo-T in the three Endo-T mutant *H. jecorina* variant (D129A, Q193A and Y195A) protein expression cultures. The gel was run and stained as described above ("*H. jecorina* growth cultures") but made with different gel ingredients: Separating gel: 15% AA:bisAA(30:1) + 7.01M glycerol + 375mM Tris-HCl pH 8.8 + 0.1% SDS + 43.8mM APS + 8.6mM TEMED; Stacking gel: 4% AA:bisAA(30:1) + 125mM Tris-HCl pH 6.8 + 0.1% SDS + 21.9mM APS + 3.44mM TEMED. The bands corresponding to the size of Endo-T were analyzed and identified with mass spectrometry (Ultraflex MALDI TOF/TOF, Bruker Daltonics) by Åke Engström at the Department of medical biochemistry and microbiology,

Biomedical centre, Uppsala, Sweden (personal communications).

MonoQ an-ion exchange chromatography
The Avicel-purified supernatant of the mutant with verified Endo-T expression (D129A) was buffer exchanged with 20mM Bis-Tris buffer (pH 6.5) using Vivaspin 20 concentrators (Sartorius Stedim Biotech) to reduce the conductivity to 1.5mS and titrated to pH 6.5. Endo-T was then purified by an-ion exchange chromatography (MonoQ 10/100 GL, GE Healthcare) at 4ºC using an ÄKTA Explorer™ 100 system (GE Healthcare). The MonoQ column was after equilibration with 20mM Bis-Tris buffer (pH 6.5) loaded with the Endo-T sample whereafter unbound proteins where washed away with 20mM Bis-Tris buffer (pH 6.5). Bound proteins were eluted with a linear gradient of 0-1M NaCl in 20mM Bis-Tris buffer (pH 6.5). The peak fractions containing Endo-T were identified with SDS-PAGE (15% manually produced gels described above in "*SDS-PAGE analysis of Avicel purified culture filtrate*") and pooled.

Superdex 200 size-exclusion chromatography
The pooled Endo-T peak fractions were concentrated to less than 2 ml (Vivaspin 20, Sartorius Stedim Biotech) and subsequently further purified by size-exclusion chromatography (HiLoad™ 16/60 Superdex 200™, GE Healthcare) using an ÄKTA Explorer™ 100 system (GE Healthcare). The superdex 200 column was after equilibration with 20mM NaAc (pH 5.0) loaded with the concentrated Endo-T sample and eluted with a constant flow of 20mM NaAc (pH 5.0). The peak fractions containing Endo-T were identified with SDS-PAGE (15% manually produced gels described above in "*SDS-PAGE analysis of Avicel purified culture filtrate*") and verified with mass spectrometry (Ultraflex MALDI TOF/TOF, Bruker Daltonics) by Åke Engström at the department of medical biochemistry and microbiology, Biomedical centre, Uppsala, Sweden.

Crystallization experiments
Pure wild type Endo-T sample, expressed by Genencor® and purified by I. Stals at the University of Gent, Belgium, at a concentration of 16.5 mg/ml, was used for Endo-T wild type enzyme crystallization. Two different Endo-T ligand samples were prepared as described above ("Ligand preparation and analysis") containing N-glycans or glycoaspargines at unknown concentrations. The Endo-T crystallization experiments were performed as described by former master student Jing Zhang[36]. Crystallization trials were set up using EasyXtal Tool crystallization plates (Qiagen, USA) and the hanging drop vapor diffusion method. Zink-free and low pH crystallization conditions at 20ºC with 8.3 mg/ml Endo-T, 0.1M Citric acid pH 2.4-3.0 and the precipitant Polyethylene glycol (PEG) 3350 (4-14%) or the precipitant mixture PEG 1500/PEG 8000 (4-14% of each PEG) were used to crystallize Endo-T. The oligomannosidic N-glycan and glycoaspargine ligands were both co-crystallized with the Endo-T solution and/or soaked into Endo-T crystals preceding freezing the crystals in liquid nitrogen. The co-crystallization with either of the ligands was performed by initially drying 1-2 µl highly concentrated ligand solution on the crystallization plate lids and subsequently placing the crystallization solution drop with Endo-T on top of the dried ligand. The soaking of Endo-T with either of the ligands was performed by initially drying 1-2 µl highly concentrated ligand solution on crystallization plate lids and subsequently placing crystallization solution on top of the dried ligands where protein crystals were later transferred during the crystal freezing procedure. Another approach used for soaking of Endo-T crystals was placing the cryoprotectant solution drop on top of the dried ligand which was subsequently used in the crystal freezing procedure.

Crystal soaking, dehydration & freezing The cryoprotectant (cryo) solution used prior to the instant freezing of all the Endo-T crystals, to dehydrate the outside of the protein and prevent crystal-damaging ice-formation, contained 0.1M Citric acid (pH 2.9), 10% glycerol and 35% PEG3350. Some crystals were soaked initially in a drop containing mother-well crystallization solution and an expected high concentration of ligand for about 30 min or 24 hours before being dipped into cryo solution and plunched into liquid nitrogen where the crystals were flash frozen together with the supporting loop. Other crystals were soaked in the cryo solution drop containing an expected high concentration of ligand for about 30-60 min before being plunch frozen in liquid nitrogen. The crystals were repeatedly microscopically checked for damage during transportation, soaking and dehydration procedures.

Crystal data collection and process The X-ray diffraction data sets from the protein crystals were obtained by the use of the high intensity X-ray providing synchrotron, beamline I911-2 at the Swedish synchrotron radiation laboratory MAX-lab at Lund University, Sweden. An X-ray source located at the home lab (Dept. of Molecular Biology, Swedish University of Agricultural Sciences, Uppsala, Sweden) equipped with a rotating anode generator (Rigaku Rotaflex RTP300) and a Mar345 IP detector, was also used for minor X-ray diffraction experiments. The collected X-ray diffraction data sets were integrated with iMosflm 1.0.0 [46]. Subsequent data processing was performed using the CCP4 program suit[47].

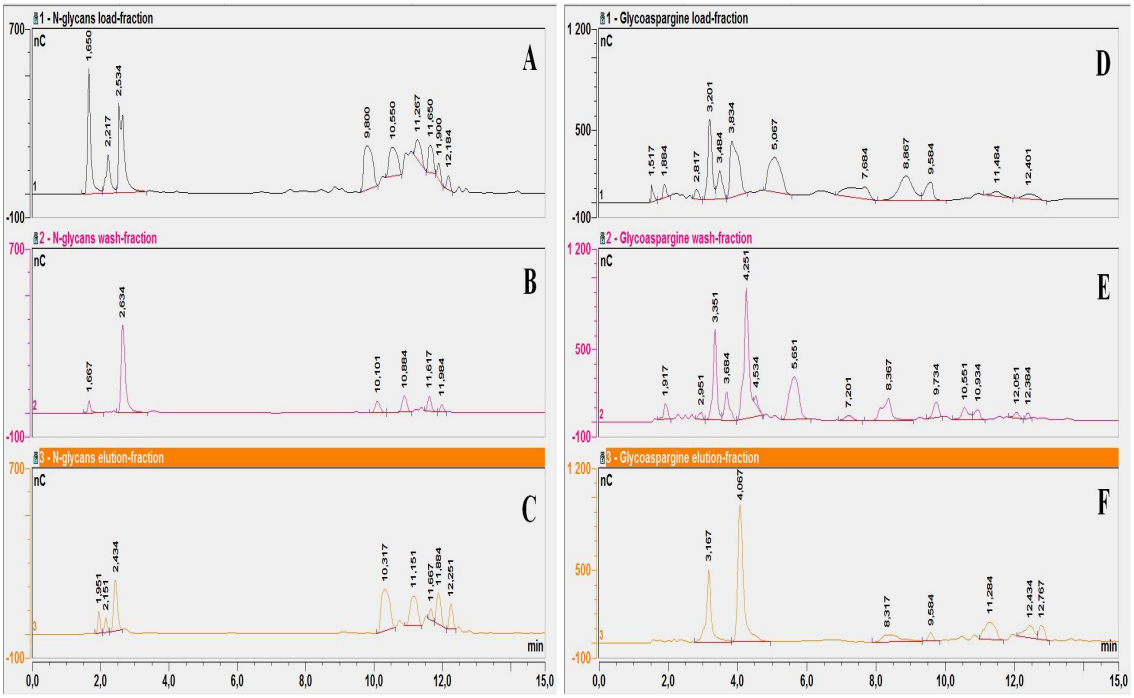


Figure 4. **HPAEC-PAD analysis of the purification of oligomannosidic N-glycans and glycoaspargines.** Chromatograms of samples containing different fractions collected during the purification of oligomannosidic N-glycans (Figure 3A-C) and glycoaspargines (Figure 3D-F) with the carbograph column (Alltech™ Extract Clean™ SPE Carbo 150mg/4ml) analyzed by high performance an-ion exchange chromatography with pulsed amperometric detection (HPAEC-PAD; ICS-3000, Dionex). The collected fractions include load (Figure 3A and D), wash (Figure 3B and E) and elution (Figure 3C and F) fractions from both types of oligomannosidic ligands.

The data was scaled and merged with Scala[48], while the structure was solved by molecular replacement using the apo-structure of Endo-T as search model using the program Phaser[44]. Refinement of the structural model was performed using Refmac 5.5 [49] and the structure model building and validating freeware program Crystallographic Object-Oriented Toolkit, or simply Coot[50]. The protein structure model visualizing program PyMOL[51] was used to envision various elements of the enzyme structure in a cartooned format.

Results

Preparation of oligomannosidic N-glycan and glycoaspargine ligands

High mannose type N-glycans were obtained from the glycoprotein RNase B and are natural ligands to Endo-T containing between 5-9 mannose residues[24]. The N-glycans were cleaved of in two different ways:a) Precise hydrolysis of the glycosidic bond between the proximal N-acetylglucosamine and the aglycone aspargine catalyzed by Peptide-N4-(acetyl-ß-glucosaminyl)-asparagine amidase F (PNGase F) or b) Pronase digestion of the whole protein molecule of RNase B into single amino acids leaving the N-glycans attached to a single aspargine residue. As the pronase digestion often is incomplete, the amount of amino acid residues attached to the N-glycans can vary, however they will despite this fact in this study be referred to as glycoaspargines. The HPAEC-PAD chromatogram of the elution-fraction from the N-glycans without aspargine (Figure 4C) displays 5 peaks probably corresponding to the 5 N-glycans containing 5-9 mannose residues and 3 early peaks probably corresponding to the 4.8M acetonitrile used to elute the bound N-glycans.

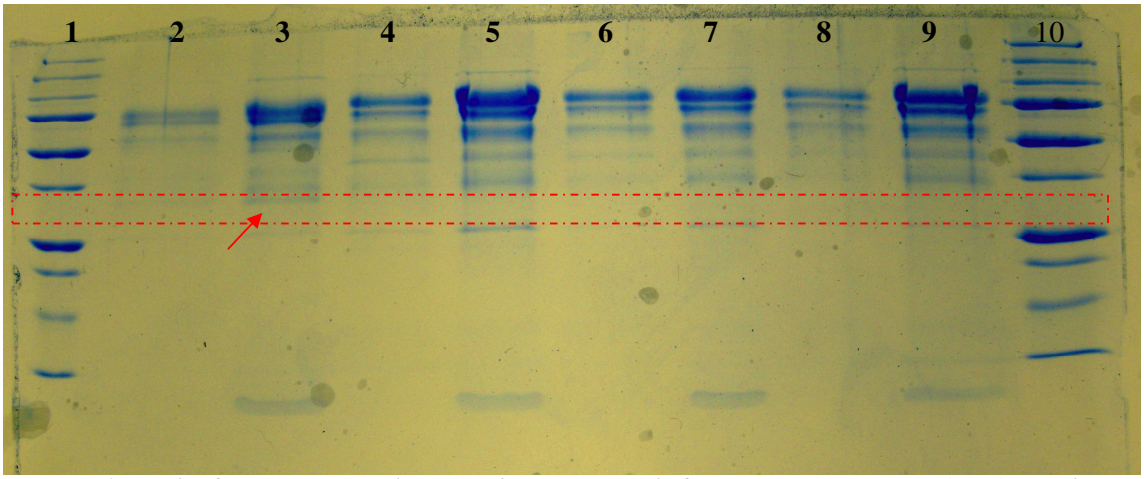


Figure 5. **Analysis of Endo-T expression levels in culture media from Endo-T mutant *H. jecorina* variants (D129A, Q193A and Y195A) and an Endo-T knock out strain.** 12% SDS-PAGE analysis of the amounts of secreted Endo-T in extracellular culture medium from three differently catalytically mutated Endo-T producing *H. jecorina* clones (D129A, Q193A and Y195A) and an Endo-T knock out strain grown in 300ml liquid 2% lactose induction media cultures. The position on the gel approximately corresponding to the size of Endo-T is indicated with a red dotted frame. A band at a size corresponding to Endo-T is indicated with red arrow. Lane 1: Low molecular weight (LMW) marker; Lane 2: Endo-T D129A *H. jecorina* culture media; Lane 3: Endo-T variant D129A *H. jecorina* culture media concentrated 20x; Lane 4: Endo-T Q193A *H. jecorina* culture media; Lane 5: Endo-T variant Q193A *H. jecorina* culture media concentrated 20x; Lane 6: Endo-T Y195A *H. jecorina* culture media; Lane 7: Endo-T variant Y195A *H. jecorina* culture media concentrated 20x; Lane 8: Endo-T Knock out *H. jecorina* culture media; Lane 9: Endo-T variant Knock out *H. jecorina* culture media concentrated 20x; Lane 10: LMW marker.

In the chromatograms from the load- and wash-fractions (Figure 4A-B), there are peaks with the same retention time as the probable N-glycans in the elution-fraction (Figure 4C) indicating loss of N-glycans in these chromatography steps. In the HPAEC-PAD chromatogram of the elution-fraction from the glycoaspargine ligands (Figure 4F) there are as in the elution-fraction of the other type of ligand (Figure 4C) 5 peaks probably corresponding to the glycoaspargines containing 5-9 mannose residues and 2 early big peaks most likely corresponding to the 4.8M acetonitrile. In the chromatograms from the load- and wash-fractions (Figure 4D-E) from the glycoaspargines there are several peaks spanning from the probable acetonitrile peaks to the probable glycoaspargine peaks. These could correspond to amino acids or peptides not precipitated during the protein precipitation, still remaining in the sample before carbograph purification.

Expression of three mutant *H. jecorina* Endo-T variants (D129A, Q193A and Y195A) and the knock out strain

The Endo-T expression levels were analyzed in extracellular culture medium from cultures of three *H. jecorina* Endo-T expressing variants (D129A, Q193A and Y195A) and an Endo-T knock out strain by SDS-PAGE analysis using 12% manually produced gels (Figure 5). Endo-T was expected to be overexpressed in the Endo-T mutant transformed *H.* jecorina clones as a lactose-induced promoter was put in front of the mutated Endo-T gene sequence coded in the transformed DNA molecule. However, no band was found at the approximate size of Endo-T (pure Endo-T sample comparison with degraded LMW marker in Figure 6A, Lane 1, 2, 7 and 8,

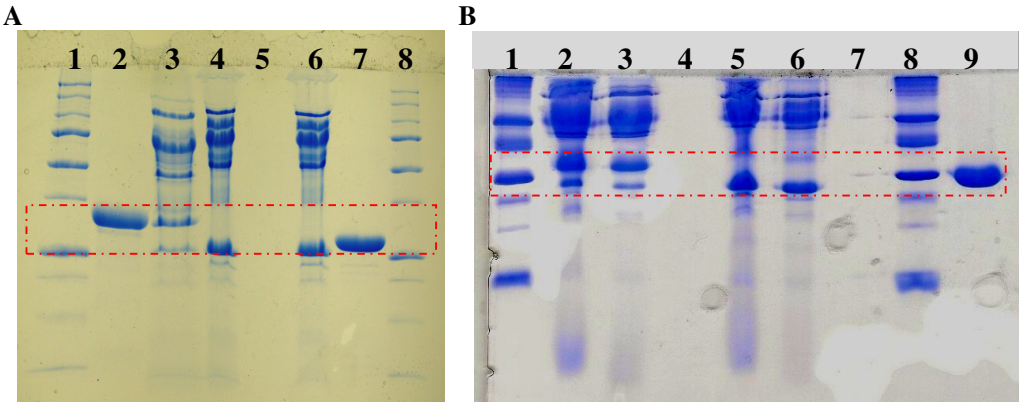


Figure 6. **Analysis of Endo-T expression levels in Avicel purified culture media from Endo-T mutant *H. jecorina* variants D129A, Q193A and Y195A. (A)** 12% SDS-PAGE analysis of the amounts of Endo-T in Avicel purified culture media samples from variant D129A and Q193A Endo-T producing *H. jecorina* clones. Lane 1: Low molecular weight (LMW) marker; Lane 2: Pure Endo-T overexpressed in *Pichia pastoris*; Lane 3: Endo-T variant D129A *H. jecorina* Avicel purified sample concentrated 30x; Lane 4: Endo-T variant Q193A *H. jecorina* expressed, Avicel purified sample concentrated 30x; Lane 5: Ultrafiltration flowthrough sample; Lane 6: Endo-T variant Q193A *H. jecorina* expressed, Avicel purified sample accidentally dried; Lane 7: Pure wild type Endo-T purified from *H. jecorina* diluted 10x; Lane 8: Low molecular weight (LMW) marker. **(B)** 15% SDS-PAGE analysis of the amounts of Endo-T in Avicel purified culture media samples from variants D129A and Y195A Endo-T producing *H. jecorina* clones. Lane 1: Low molecular weight (LMW) marker; Lane 2: Endo-T variant D129A *H. jecorina* expressed, Avicel purified sample concentrated 120x; Lane 3: Endo-T variant D129A *H. jecorina* expressed, Avicel purified sample concentrated 60x; Lane 4: Empty; Lane 5: Endo-T variant Y195A *H. jecorina* expressed, Avicel purified sample concentrated 120x; Lane 6: Endo-T variant Y195A *H. jecorina* expressed, Avicel purified sample concentrated 60x; Lane 7: Empty; Lane 8: Low molecular weight (LMW) marker; Lane 9: Pure wild type Endo-T purified from *H. jecorina* diluted 50x.

area indicated by red dotted frame in Figure 5) in either of the culture samples lanes indicating Endo-T overexpression. There was only one single band at a size equivalent to the size of Endo, a faint band from the Endo-T mutant D129A *H. jecorina* culture medium (Figure 5, Lane 3, band indicated with red arrow) but the result was not conclusive.

Endo-T purification

After being harvested and filtered (to get rid of fungal mycelia), the culture media were purified by adsorbing the cellulases, having a cellulose binding module, onto Avicel and then pelleting and discarding the Avicel with bound protein together. Following the enzyme adsorption onto Avicel, the different Endo-T mutant samples were analyzed with 12% and 15% SDS-PAGE gels. A few bands from all three Endo-T mutant samples were located at sizes spanning from the size of the pure Endo-T sample expressed in *Pichia pastoris* to the pure Endo-T sample

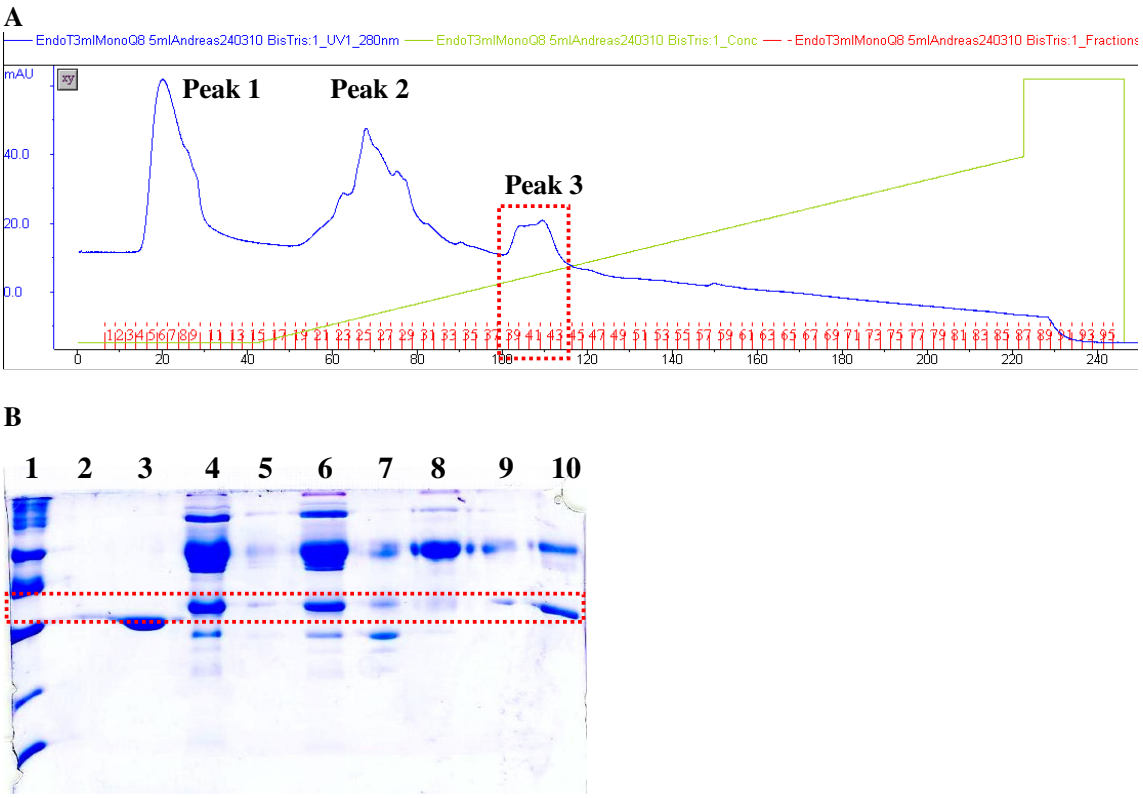


Figure 7. **Purification of Endo-T D129A mutant by an-ion exchange chromatography.** An-ion exchange chromatography purification of Endo-T D129A mutant with a MonoQ 10/100 GL column (GE Healthcare) in 20mM Bis-Tris (pH 6.5) buffer and eluted with a linear gradient of 1M NaCl. Samples were collected in 2.5 ml fractions. **(A)** Chromatogram of collected MonoQ-purified fractions (red) showing protein absorbance at 280nm (blue) and linear 0-1M NaCl gradient (green). The 3rd peak corresponding to Endo-T is highlighted with a red dotted frame **(B)** 15% SDS-PAGE analysis of MonoQ-separated peak fractions. Endo-T bands are highlighted by a red dotted frame. Lane 1: Low molecular weight (LMW) marker; Lane 2: Empty; Lane 3: Pure wild type Endo-T purified from *H. jecorina* diluted 50x; Lane 4: Endo-T variant D129A *H. jecorina* expressed, Avicel purified sample concentrated 60x; Lane 5: Empty; Lane 6: Rescued sample from unsuccessful MonoQ separation with 20mM NaAc (pH 5.0) buffer; Lane 7: MonoQ fraction tube 6 from 1st peak; Lane 8: MonoQ fraction tube 25 from 2nd peak; Lane 9: Empty (well contamination from lane 8); Lane 10: MonoQ fraction tube 40 from 3rd peak:

expressed in *H. jecorina* (Figure 6A and 6B). These gel bands from Endo-T mutant D129A and Y195A samples (Figure 6B, bands inside the red dotted frame) were analyzed by mass spectrometry and one analyzed band from the Endo-T mutant D129A sample was identified as being Endo-T (data not shown). After verifying the presence of Endo-T in one of the samples, an an-ion exchange chromatography purification step with a MonoQ 10/100 GL column (GE Healthcare) was utilized to purify the deglycosylating protein. Three major peaks were acquired, one unbound fraction (1st peak), one loosely bound fraction (2nd peak) and one strongly bound fraction (3rd peak) (Figure 7A). A top peak fraction from all the peaks was analyzed by 15% SDS-PAGE (Figure 7B). A strong Endo-T band was found in the sample from the 3rd peak (Figure 7B, lane 10) and a faint Endo-T band was seen in the unbound fraction sample from the 1st peak (Figure 7B, lane 7). The third peak was therefore collected and used for further purification.

The Avicel and MonoQ-purified Endo-T D129A mutant sample was subsequently

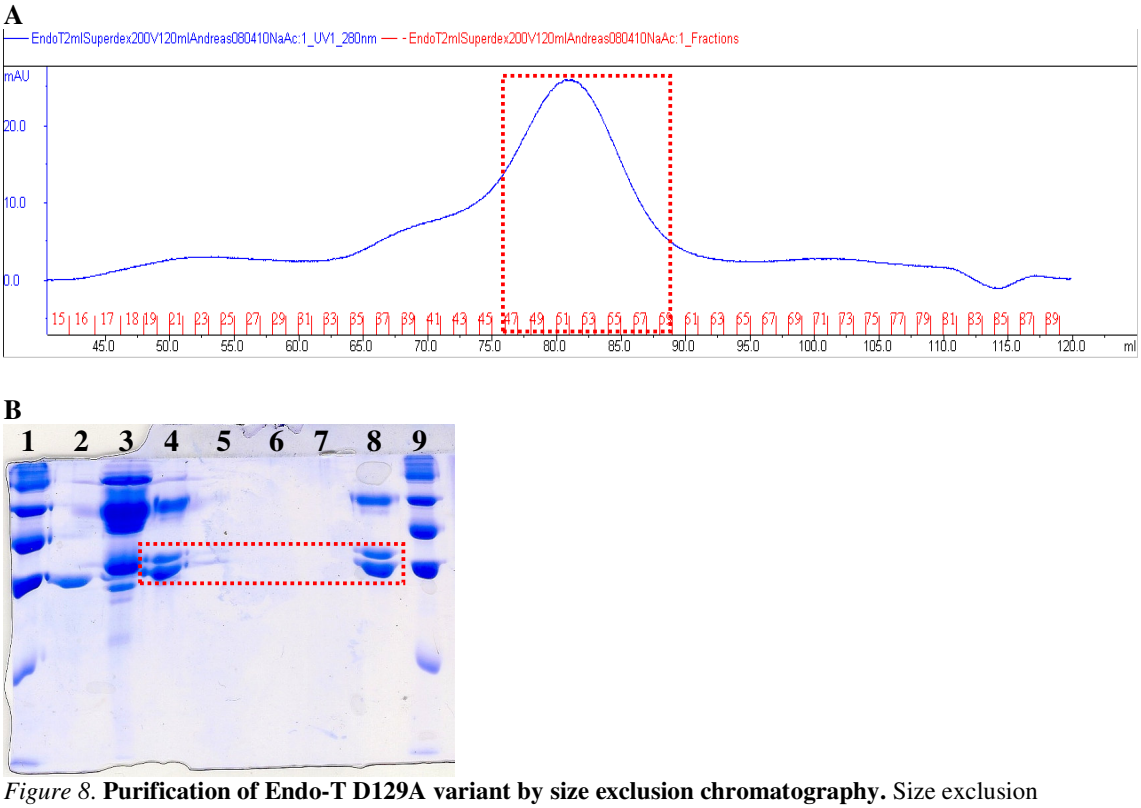


Figure 8. **Purification of Endo-T D129A variant by size exclusion chromatography.** Size exclusion chromatography of Endo-T D129A variant with a HiLoad 16/60 Superdex 200 column (GE Healthcare) in 20mM NaAc (pH 5.0) buffer. Samples were collected in 1 ml fractions. **(A)** Chromatogram of collected Superdex 200-purified fractions (red) showing protein absorbance at 280nm (blue). Collected peak fractions are highlighted with a dotted red frame. **(B)** 15% SDS-PAGE analysis of elution fractions from the Superdex 200 separated peak. Endo-T bands are highlighted with a red dotted frame. Lane 1: Low molecular weight (LMW) marker; Lane 2: Pure wild type Endo-T purified from *H. jecorina* diluted 100x; Lane 3: Endo-T variant D129A *H. jecorina* expressed, Avicel purified sample concentrated 60x; Lane 4: Pooled Superdex 200 peak fractions 47-59 concentrated 4x; Lane 5-7: Empty; Lane 8: Pooled Superdex 200 peak fractions 47-59 concentrated 4x; Lane 9: Low molecular weight (LMW) marker.

further purified by size exclusion chromatography using a HiLoad 16/60 Superdex 200 (GE Healthcare) column. A double peak (Figure 8A) was obtained where the last major part of the peak (peak fraction 47-59) was analyzed by 15% SDS-PAGE (Figure 8B, lane 4 and 8) which demonstrated the existence of 3 strong separate bands. Two of the bands were present at the size of Endo-T while the third band was present at a larger size. The two bands present at a smaller size were analyzed by mass spectrometry and both of these bands matched the spectra of the wild type Endo-T (data not shown).

Endo-T crystallization

Endo-T was co-crystallized and/or soaked with N-glycans and glycoaspargines, natural ligands to Endo-T[24], to produce ligand bound Endo-T crystals. The crystallization experiments were performed at low pH (2.4-3.0) in an attempt to inactivate the catalytic activity of Endo-T and thus avoid hydrolysis of the ligand when bound to Endo-T in the crystal structure. Many different crystallization conditions were found to successfully give rise to Endo-T crystals and two optimal conditions for Endo-T crystal growth were found from which crystals were used for X-ray diffraction data collection: 1) 8.3 mg/ml Endo-T mixed with equal amounts of 0.1M Citric acid pH 2.4-3.0 and 4-6% PEG 3350; and 2) 8.3 mg/ml Endo-T mixed with equal amounts of 0.1M Citric acid pH 2.4-3.0, 4-6% PEG 1500 and 4-6% PEG 8000. The crystals grown in these conditions, both without and co-crystallized with ligand, displayed mixed morphology. The crystal shapes spanned from rod structures to cubic and bar structures both big and small (Figure 9).

X-ray diffraction and processing

X-ray diffraction data sets from Endo-T were collected at the Swedish synchrotron radiation laboratory Max-lab at Lund University, Lund, Sweden. Data were successfully collected from 5 Endo-T-crystals grown and soaked at 3 different conditions: I) Co-crystallization with 2 µl dried oligomannosidic N-glycans (unknown concentration) in 0.1M Citric acid pH 2.4 and 4-6% PEG3350 cryoprotected in cryo containing 1 µl dried oligomannosidic N-glycans (unknown concentration), II) Co-crystallization with 1 µl dried oligomannosidic glycoaspargines (unknown concentration)

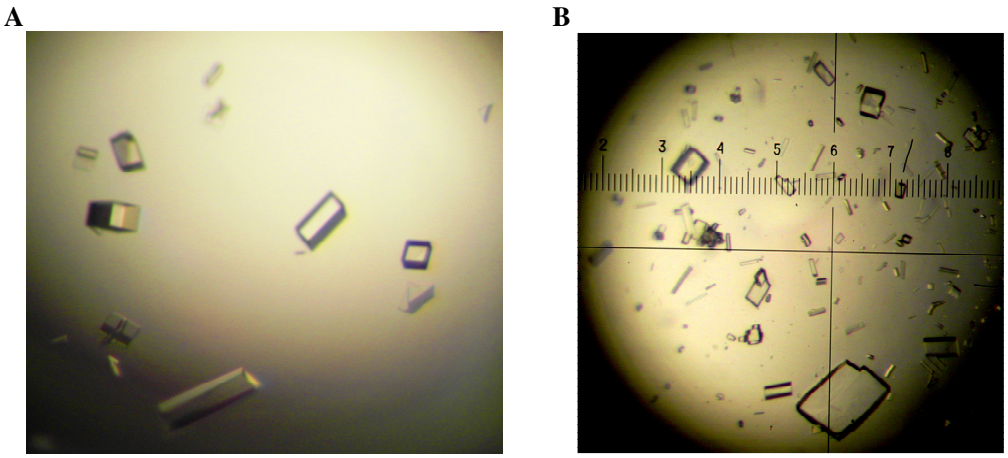


Figure 9. **Endo-T crystallization.** 8.3 mg/ml pure Endo-T sample were crystallized by the hanging drop vapor diffusion method mixing equal amounts of 0.1M Citric acid pH 2.4-3.0 and 4-6% of different PEGs. **(A)** Endo-T crystals grown with 0.1M Citric acid pH 2.8 and 6% PEG 3350. Crystals emerged after a week. **(B)** Endo-T crystals grown with 0.1M Citric acid pH 2.4, 4% PEG 1500 and 4% PEG 8000 and co-crystallized with 1ul dried oligomannosidic N-glycans of unknown concentration. Crystals emerged after 24 hours.

in 0.1M Citric acid pH 2.4, 4-6% PEG1500 and 4-6% PEG8000 cryoprotected in cryo containing 2 µl dried oligomannosidic glycoaspargines (unknown concentration) and III) Crystallization in 0.1M Citric acid pH 2.8 and 6% PEG3350 soaked in cryo containing 2 µl dried oligomannosidic glycoaspargines (unknown concentration) for about 30 min. Clear electron density was observed for a bound ligand in every co-crystallized Endo T-ligand complex structure that was solved. The data that resulted into the best structure model was obtained from a crystal grown and soaked according to the first conditions. The structure has been refined to 1.65Å resolution with a final R- and R_{free} value of 0.1871 and 0.2142, respectively.

The ligand that was found bound in the active site of Endo-T consists of a processed and incomplete oligomannosidic N-glycan Manα1-6(Manα1-3)Manα1-6Manβ1-4GlcNAc with missing electron density from the pentasaccharide core C3 linked oligomannosidic branch (Figure 10). The overall structure of the crystallized Endo-T is an $(\beta/\alpha)_8$-barrel fold with eight β-strand/loop/α-helix units where the catalytic site is positioned on top of the core of the barrel-structure (Figure 11). Two single N-acetylglucosamine units are found bound to the protein in the structure, one bound to Asn70 and the other to Asn240. The introduced ligand is found bound in the substrate binding cleft at the top of the β-barrel. The distal glycone part of the N-glycan is bound over the shallow loop of $(\beta/\alpha)_8$-barrel unit 2, while the proximal N-acetylglucosamine is found in the center on top of the core of the β-barrel at the catalytic site near the catalytic residues Asp129 and Glu131 and the stabilizing residues Tyr195 and Trp259 (Figure 12). The distance between the outer carbon of the side chain of Glu131 and the anomeric carbon of the hydrolyzed N-acetylglucosamine is 6.88Å (Figure 12). Electron density which appear to correspond to a furanose ring was found in a close approximate to the bound ligand on top of the shallow barrel loop unit 8 (Figure 10 and 14).

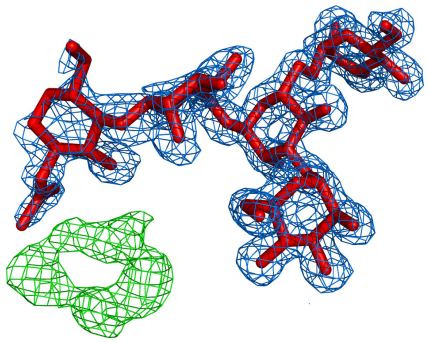


Figure 10. **Electron density maps for a (GlcNac)-Man_4 and an unidentified molecule.** Cartoon visualization of the substrate electron density map (marine blue; sigma level 1) fitted with Manα1-6(Manα1-3)Manα1-6Manβ1-4GlcNAc (red) and an unidentified furanose shaped electron density map (green; sigma level 1).

Discussion

Until just recently there was no knowledge about eukaryote ENGase activity among the glycoside hydrolases of family 18 (GH18)[16]. A peculiar amount of single N-acetylglucosamine residues, found on N-glycosylation sites of *H. jecorina*-secreted cellulases, prompted the idea of the existence of endo-β-*N*-acetylglucosaminidase activity in this ascomycete[37-41]. The hypothesis led to the discovery of a fungal Endo-β-*N*-acetylglucosaminidase, with the chitinase active site motif of the ENGases of glycoside hydrolase family 18, named Endo-T after the former name of the fungus which the enzyme originates, *Trichoderma reesei*[24]. Simultaneously, on the other side of the planet, another Endo-β-*N*-acetylglucosaminidase was found expressed by a basidiomycete, *Flammulina velutipes*, named Endo-FV. In these studies, several additional species of ascomycetes and basidiomycetes were also found to encompass this activity[24, 25].

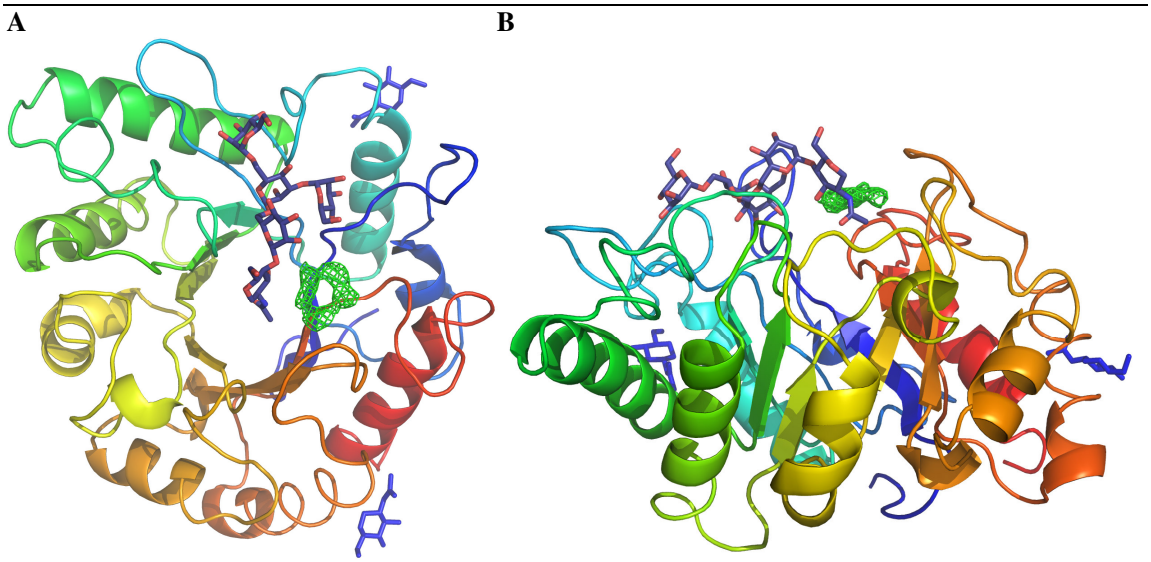


Figure 11. **Endo-T with bound processed substrate (GlcNac)-Man$_4$.** Cartoon visualization of Endo-T (rainbow colored carbon backbone from blue at the N-terminous to red at the C-terminous of the protein) with a bound (GlcNac)-Man$_4$ (blue/red). Two bound single N-acetylglucosamines are shown linked to Asn70 and Asn240 (blue). Furanose-shaped electron density is shown in green. **(A)** Top view of the $(\beta/\alpha)_8$-barrel structure fold with the substrate bound on top of the central β-barrel. **(B)** Side view of the $(\beta/\alpha)_8$-barrel structure fold with the substrate bound on top of the central β-barrel.

Except the urge to learn and understand more about the action of these enzymes there is also a practical demand for this type of activity. The usage of Endo-H found to be secreted by *Streptomyces plicatus* is vast in the field of glycoproteomics and also X-ray crystallography[21, 26, 27]. The high-price of the enzyme has led to the idea of using Endo-T as a competing alternative to Endo-H usage as it comprise similar substrate specificities. As of now, a great deal has been elucidated about the enzyme, but there are still a lot left to be investigated, such as the activating mechanisms of N- and C-terminal processing[24], possible existence of transglycosylation properties and the mechanisms behind the substrate specificity. The latter was the driving force behind this study where the goal was to introduce a ligand in the active site of crystallized Endo-T under such circumstances that it would not be processed and/or released to enable structure determination of Endo-T with its natural substrate bound in the active site, unprocessed. To render this possible, two angles of attack were chosen. The first one, where the key-amino acids, involved in the catalytic activity, were replaced by inactive alanine residues and the second one, where extreme environmental conditions in the form of extreme pH were used to disturb the energy balances in the different steps of the catalytic reaction. The first approach was unfortunately derailed during the bioengineering of highly Endo-T-mutant expressing *H. jecorina* clones due to miscommunications and only one low-expressing variant were acquired resulting in a low concentration sample of mutant Endo-T with rather poor purity. The other approach was partially successful, yielding Endo-T structures with the substrate bound in the active site. However, the substrate was processed and without adequate electron density information about several mannose residues in the distal part of the oligomannosidic N-glycan, including the whole oligomannosidic branch linked to

C3 in one of the pentasaccharide mannose residues (Figure 10).

Substrate binding interactions with Endo-T

Endo-T binds its substrate (usually attached to proteins) with the top of the $(\beta/\alpha)_8$-barrel (indicated by the upward direction of the β-barrel β-sheets) parallel to the substrate glycans chain (Figure 11B). The different parts of the N-glycan are nicely fitted into low energy pockets (Figure 13). The tight and high loops of unit 6 and 7 leads to the conclusion that the substrate linked aglycone docks above loop unit 5 where the side chains of the amino acids are pointed down or parallel to the $(\beta/\alpha)_8$-barrel top plain (Figure 11B and 13B). Hydrophobic residues Leu171 of unit 5 and Phe198 of unit 6 are positioned at the probable docking area of the aglycone and are probably stabilizing the Endo-T-aglycone interaction (Figure 13, indicated by black). The proposed catalytic amino acid residues are positioned in the area beneath the N-acetylglucosamine of the bound substrate (Figure 12 and 13(pink residues)). The proposed stabilizing residues Tyr195 and Trp259 are indeed located around the N-acetyl arm of the N-acetylglucosamine and seem to stabilize the substrate-Endo-T interaction by keeping the N-acetyl arm in a hydrophobic environment. The proposed catalytic residues Asp129 and Glu131 are situated 9.8 and 6.9 Å beneath the anomeric carbon of the substrate N-acetylglucosamine, respectively. In a plausible reaction event the two N-acetylglucosamines of the unprocessed substrate could descend closer to these residues to facilitate the hydrolysis of the substrate. At the distal glycone part of the bound substrate, including four mannose residues (up to nine in the unprocessed substrate), in the area of loop unit 2, the branched carbohydrate structure is fitted in a highly specific polar cleft demonstrating the substrate specificity of Endo-T (Figure 13, indicated by brown). These residues of the glycone cleft, including mostly aspargine residues located on loop units 1 and 2, are to a high degree conserved among the ENGases of glycoside hydrolases family 18 [42]. The additional mannose residues, extended

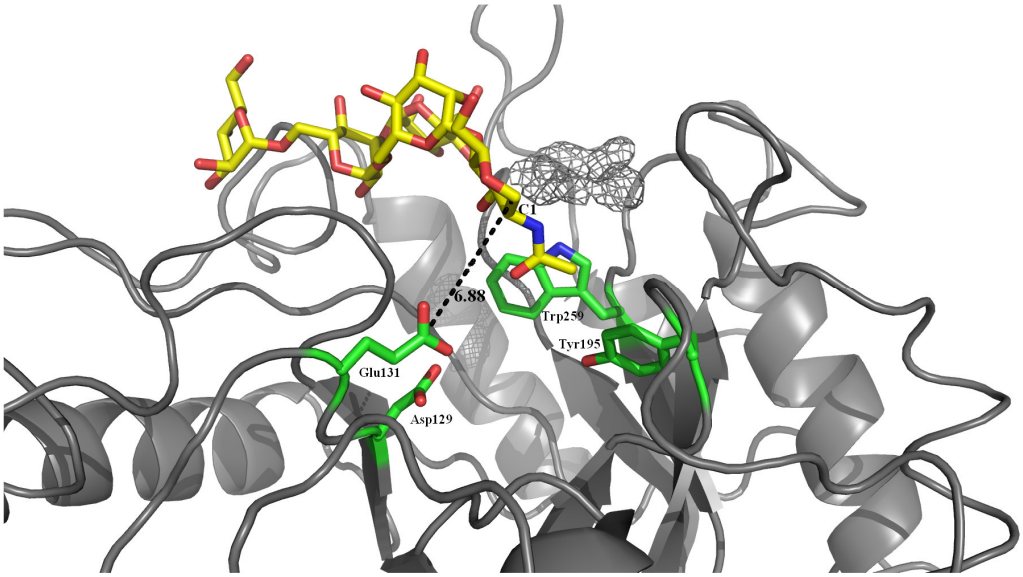


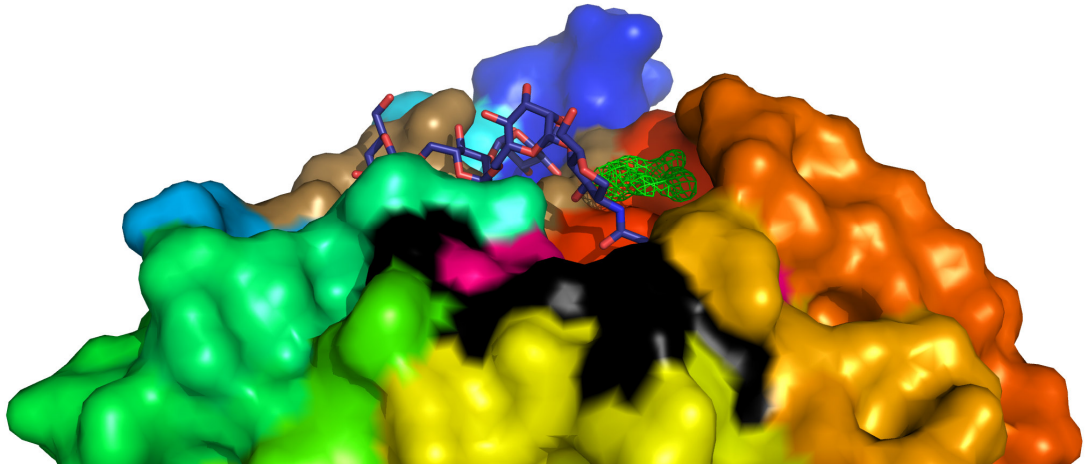
Figure 12. **Active site of Endo-T with bound (GlcNac)-Man$_4$.** Cartoon visualization of the active site of Endo-T (grey) with the bound (GlcNac)-Man$_4$ (yellow/red). Catalytic (Asp129 and Glu131) and stabilizing (Tyr195 and Trp259) residues are highlighted in green. The distance between Glu131 and C1 of the N-acetylglucosamine is shown in black (6.88Å). Electron density of furanose is lightly displayed (grey).

from distal branched structure of the bound substrate, would be positioned slightly upwards from the $(\beta/\alpha)_8$-barrel plain and reach a position outside of the Endo-T structure (visible in figure 13B). The missing branch consisting of up to 3 mannose residues is normally linked to the C3 carbon of the N-acetylglucosamine-linked mannose residue of N-glycans. This C3 bond is directed upwards towards free space indicating that the oligomannosidic branch is not bound or loosely bound by Endo-T. This could also explain the loss of electron density information for these residues as the flexibility results in unfocused X-ray scattering.

Unidentified furanose-shaped electron density

After the last refinement, electron density appeared next to the bound ligand above a cleft formed by loop unit 8 of the enzyme (Figure 10, 11, 13 and 14).

A



B

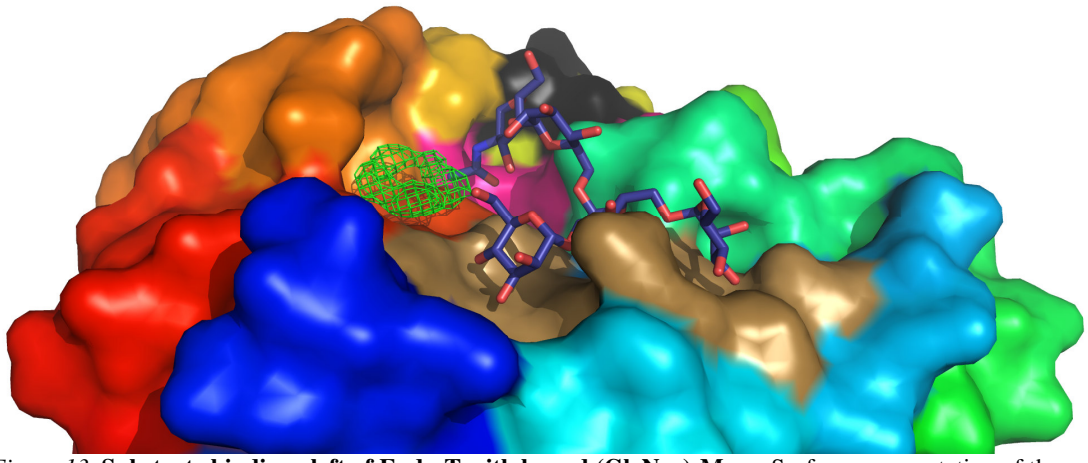


Figure 13. **Substrate binding cleft of Endo-T with bound (GlcNac)-Man$_4$.** Surface representation of the binding cleft of Endo-T. The protein residues are rainbow colored (from blue at the N-terminous to red at the C-terminous) except for the aglycone residue area (black), catalytic residue area (pink) and the distal glycone residue area (brown). The substrate (GlcNac)-Man$_4$ is shown bound in the cleft (blue/red). Furanose electron density is shown in green. **(A)** Aglycone entrance view of the binding cleft. **(B)** Distal glycone view of the binding cleft.

It has a shape of a furanose ring and fits very well in what appears to be a cleft intended for substrate binding. As the C3-branched distal mannose residue of the substrate is elongated in the C2 position, which is directed in the opposite direction to the electron density, the furanose is not likely to be part of the distal glycone part of the substrate. Hybrid type N-glycans can sometimes include an N-acetylglucosamine residue linked to C4 of the proximal mannose residue of the pentasaccharide core. However, N-acetylglucosamine is a six-membered pyranose ring and the bond of C4 where it is attached is pointing in the opposite direction to the electron density[5]. There are four polar amino acids within a distance of 4Å from the furanose density, including two glutamic acids (Glu229 and Glu260), a glutamine (Gln14) and a threonine (Thr15), which creates a positive environment for a bound carbohydrate residue. The stabilizing residue, Trp259 is also located within 3Å underneath the furanose density. All in all an oddly well constructed carbohydrate-friendly polar cleft with some unidentified molecule bound to it.

Structural comparison between Endo-T and GH18 homologues Endo-H and Endo-F3

The sequence identity between Endo-T and its $(\beta/\alpha)_8$-barrel folded glycoside family 18 homologues Endo-H from *Streptomyces plicatus*[21] and Endo-F3 from *Elizabethkingia meningoseptica*[22] is low. This fact is demonstrated by the highly dissimilar structures of the $(\beta/\alpha)_8$-barrel loops. A short glimpse of the family member proteins superimposed will also reveal differences in α-helix arrangements (Figure 15). The similarities are instead found in the residues involved in the substrate-protein interactions. The amino acids involved in substrate recognition and hydrolysis are in most cases conserved within the ENGases of GH18.

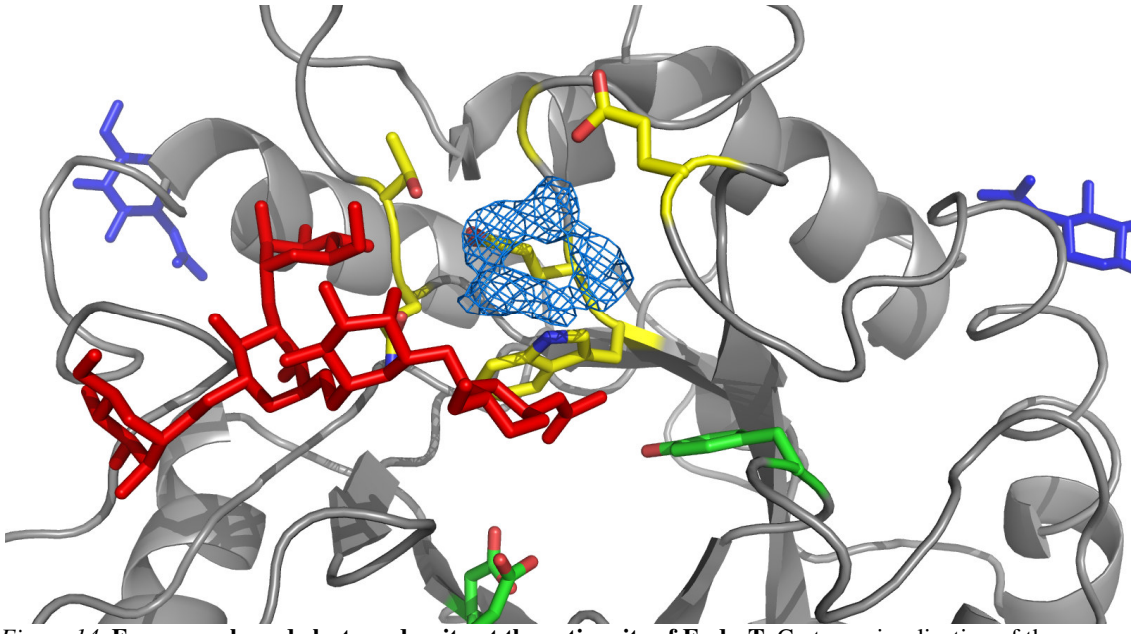


Figure 14. **Furanose-shaped electron density at the active site of Endo-T.** Cartoon visualization of the environment around a furanose-shaped electron density (marine blue). Catalytic residues (green/red), residues in the area of the furanose-shaped density (yellow/red) and the (GlcNac)-Man_4 ligand (red) are highlighted in the visualization. Two bound N-acetylglucosamines are shown linked to Asn70 and Asn240 (blue).

Endo-H has similar substrate specificities as Endo-T as both enzymes hydrolyze high-mannose type N-glycans and are unable to process complex type N-glycans[21, 24] while Endo-F3 hydrolyze bi- and triantennary complex-N-glycans[22]. The loop regions in unit 5 and 6, where the aglycone protein docks, demonstrate some differences between the homologues. Endo-H has very short and low loop regions and misses the two α-helixes of unit 5 and 6 while Endo-T and Endo-F3 possess larger and longer loops and encompass both helixes of unit 5 and 6. The loop units 5 and 6 of Endo-F3 are relatively long, creating steric hindrance to larger glycoproteins causing it to be less effective against these macromolecules[30]. Endo-T is something in between but is surely somewhat affected by loop units 5 and 6 during the docking of glycoproteins. Endo-H, on the other hand, is together with its structure homologue Endo-F1[22, 43], probably two of the most effective ENGases against large glycoproteins because of their shallow and flexible barrel units 5 and 6[42]. The catalytic aspartic acid and glutamic acid are conserved in all three homologues attached in the border between the β-sheet and loop of unit 4 but with some different positioning (not shown). The aspartic acids (Endo-T: Asp129, Endo-H: Asp 130, Endo-F3: Asp126) of the homologues are very similar in position, located within 1.4Å of each other when these structures are superimposed. The glutamic acids are slightly more segregated (Endo-T: Glu131, Endo-H: Glu132, Endo-F3: Glu128), where Endo-H is 1.8Å closer to the center of the active site compared to the other two which are very similarly positioned. These differences are small but the change of 1.8Å in glutamic acid positioning in Endo-H could result in more effective hydrolysis of substrates. Finally, the distal glycone area around loop unit 2, where there are differences. Endo-T and Endo-H have a very similar loop structures except for the fact that loop unit 2 of Endo-H includes anti-parallel β-sheets. These loops are long but very low in comparison to loop unit 2 of Endo-F3 which is compact and creates a high wall. This is demonstrated by the substrate specificity of Endo-F3 where the natural biantennary substrate is bound with

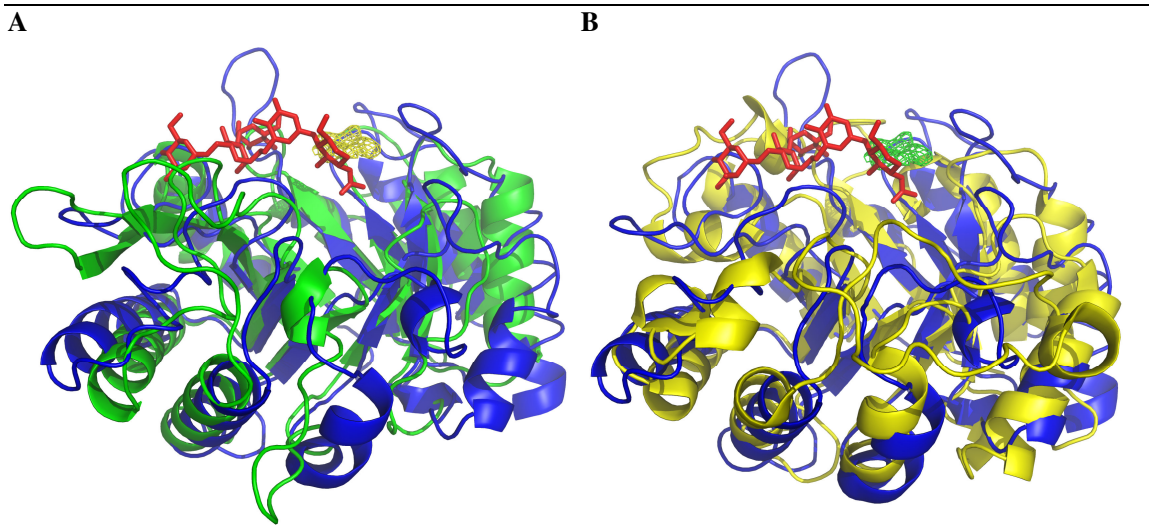


Figure 15. **Structural comparison with homologue proteins of GH18.** Cartoon visualization of Endo-T (blue) superimposed with **(A)** Endo-H (green; PDB code: 1EDT) from *Streptomyces plicatus* and **(B)** Endo-F3 (yellow; PDB code: 1EOM) from *Elizabethkingia meningoseptica*. The (GlcNac)-Man_4 ligand is shown in red together with the furanose electron density (yellow and green, respectively).

the distal part of the N-glycan directed upwards from the $(\beta/\alpha)_8$-barrel into free space and not along the top of the $(\beta/\alpha)_8$-barrel across unit 2 as for Endo-T and probably Endo-H.

Substrate binding comparison between a substrate from Endo-F3 and the bound substrate of Endo-T

An Endo-F3 structure with a biantennary complex N-glycan ligand bound in the active site (PDB code: 1EOM) was superimposed with Endo-T with the bound oligomannosidic N-glycan. The protein part of the Endo-F3 ligand complex was then removed to compare the positions of the two ligands in Endo-T (Figure 16). The anomeric carbon of the proximal N-acetylglucosamine of the complex type ligand at the active site was about 2.5Å closer to the catalytic residue, Glu131, compared to the distance between the anomeric carbon of the proximal N-acetylglucosamine of the oligomannosidic ligand and Glu131. As previously mentioned (see "Substrate binding interactions with Endo-T") the two proximal N-acetylglucosamine residues of the unprocessed high-mannose type N-glycan might descend closer to Glu131 during hydrolysis thereby enabling the reaction. The distal glycone part of the two substrates also display some differences, where the complex type substrate is wider and therefore takes up more space in the binding cleft. However, because of its distal direction upwards from the $(\beta/\alpha)_8$-barrel and into free space, it is not affected by the loop region around unit 2 as for the substrate of Endo-T. The positioning, and in this case the tilt, of the substrates can therefore explain the substrate specificities of the different homologue ENGases of glycoside hydrolases family 18.

Concluding remarks

A structure of Endo-β-*N*-acetylglucosaminidase-T was solved with a processed (GlcNac)-Man_4 bound at the active site of a fully catalytically competent wild type enzyme, by extreme

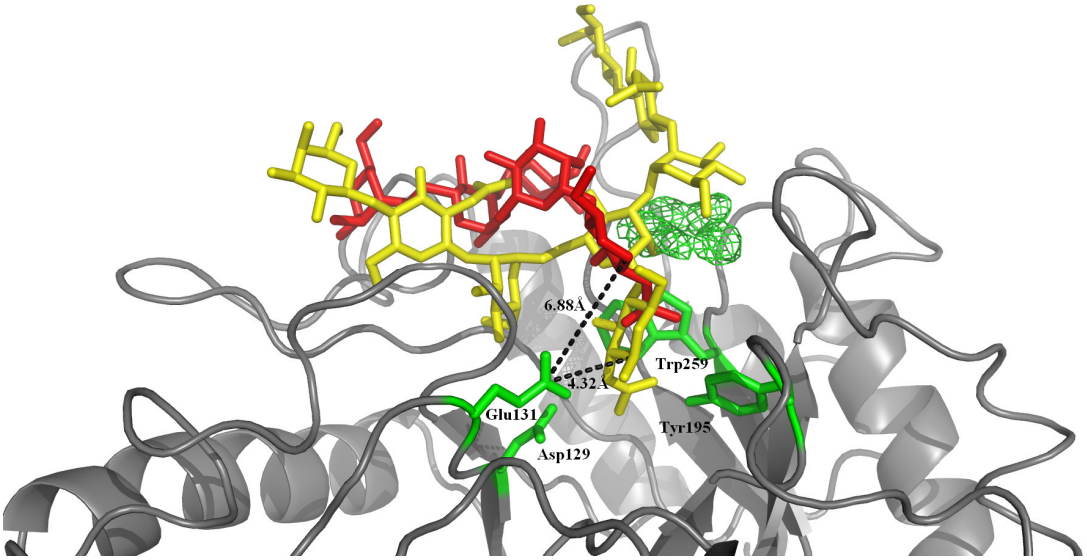


Figure 16. **Active site of Endo-T with bound (GlcNac)-Man_4 and superpositioned Endo-F3 biantennary octasaccharide ligand.** Cartoon visualization of the active site of Endo-T (grey) with a bound (GlcNac)-Man_4 (red) superimposed with the Endo-F3 active site-positioned biantennary octasaccharide ligand (yellow; PDB code: 1EOM). Catalytic (Asp129 and Glu131) and stabilizing (Tyr195 and Trp259) residues together with furanose electron density is shown in green.

pH-inactivation. The anomeric carbon of the proximal N-acetylglucosamine of the substrate was positioned 6.88Å from the suggested catalytic amino acid residue Glu131, indicating a descending motion of the two N-acetylglucosamines of the unprocessed oligomannosidic N-glycan substrate during hydrolysis. The distal part of the glycone was precisely fitted in a polar cleft of mostly aspargines around loop unit 2. The position of the proximal part of the substrate and the loop-structure of Endo-T suggests aglycone docking centered over $(\beta/\alpha)_8$-barrel unit 5. Future prospects include further inactivation of Endo-T by lowering the pH or by site-directed mutagenesis on catalytic amino acid residues. An engineered N-glycan substrate has also been produced that can not be hydrolyzed by Endo-T and will therefore bind in the active site of Endo-T and stay there through crystallization and freezing processes. The result will be un-hydrolyzed substrates bound in the active site supplying information about both aglycone and glycone interactions with Endo-T. Some method must also be proposed for stabilizing the distal oligomannosidic part of the N-glycan substrates when bound in Endo-T crystals to enable the collection of electron density information about these glycan residues. Exploration of these interactions between ENGases and their substrates will enlighten the purpose for the existence of these carbohydrate active enzymes and facilitate the manufacturing of programmed glycoside hydrolases, built for a purpose in different fields, such as biomass conversion or glycotherpeutics.

Acknowledgements

Thanks are in order to several people who helped me along the road of enlightenment. First of all I would like to thank my supervisor Mats Sandgren for taking me in under his wings, for helping me along the way and teaching me about both research and future perspectives. I would like to thank Ingeborg Stals for her cheerful guidance and for flying me over to Gent and letting me learn about both their amazing experimental methods and Belgium culture. I would also like to thank Saeid Karkehabadi for helping me with protein purification, the crystallization and driving me from Lund where he taught me about and showed me how to collect X-ray diffraction data and a lot more, Jerry Ståhlberg for guiding me through Mats-less times and teaching me glycomics, Nils Mikkelsen for happily helping me with the processing of the diffraction data and model building, Henrik Hansson for helping me with the sugar analysis and protein visualization, Åke Engström for analyzing my protein gel samples with MS, Roy Saumendra Prasad, Dieter Depuydt, Nathalie Bouly, Robert Olson and Jurgen Van Impe for helpful tips and happy social moments. Finally, I would like to thank all people of the department of molecular biology, Swedish University of Agricultural Science, Uppsala, Sweden for your help and for creating a joyful working atmosphere that made this project work, solely a pleasure.

References

1. A Varki, R. Cummings, J. Esko, H. Freeze, G. Hart, and J. Marth, *Essentials of Glycobiology*, 1st ed (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, 1999)
2. P Sears, and C H Wong, 'Enzyme action in glycoprotein synthesis', *Cellular and Molecular Life Sciences: CMLS*, 54 (1998), 223-252.
3. A Varki, 'Biological roles of oligosaccharides: all of the theories are correct', *Glycobiology*, 3 (1993), 97-130.
4. R A Laine, 'A calculation of all possible oligosaccharide isomers both branched and linear yields 1.05 x 10(12) structures for a reducing hexasaccharide: the Isomer Barrier to development of single-method saccharide sequencing or synthesis systems', *Glycobiology*, 4 (1994), 759-767.
5. A Kobata, 'Structures and functions of the sugar chains of glycoproteins', *European*

Journal of Biochemistry / FEBS, 209 (1992), 483-501.

6. D E Koshland, 'Stereochemistry and the mechanism of enzymatic reactions', *Biol. Rev. Camb. Philos. Soc* 28 (1953), 416–436.

7. M L Sinnott, 'Catalytic mechanisms of enzymic glycosyl transfer', *Chem. Rev* 90 (1990), 1171–1202.

8. J D McCarter and S.G. Withers, 'Mechanisms of enzymatic glycoside hydrolysis', *Curr. Opin. Struct. Biol* 4 (1994), 885–892.

9. B Henrissat, 'A classification of glycosyl hydrolases based on amino acid sequence similarities', *The Biochemical Journal*, 280 Pt 2 (1991), 309-316.

10. C Chothia, and A M Lesk, 'The relation between the divergence of sequence and structure in proteins', *The EMBO Journal*, 5 (1986), 823-826.

11. B Henrissat, M Claeyssens, P Tomme, L Lemesle, and J P Mornon, 'Cellulase families revealed by hydrophobic cluster analysis', *Gene*, 81 (1989), 83-95.

12. B Henrissat, 'Weak sequence homologies among chitinases detected by clustering analysis', *Protein Sequences & Data Analysis*, 3 (1990), 523-526.

13. B Svensson, 'Regional distant sequence homology between amylases, alpha-glucosidases and transglucanosylases', *FEBS Letters*, 230 (1988), 72-76.

14. E A MacGregor, and B Svensson, 'A super-secondary structure predicted to be common to several alpha-1,4-D-glucan-cleaving enzymes', *The Biochemical Journal*, 259 (1989), 145-152.

15. E Raimbaud, A Buleon, S Perez, och B Henrissat, 'Hydrophobic cluster analysis of the primary sequences of alpha-amylases', *International Journal of Biological Macromolecules*, 11 (1989), 217-225.

16. B L Cantarel, P M Coutinho, C Rancurel, T Bernard, V Lombard, B Henrissat, 'The Carbohydrate-Active EnZymes database (CAZy): an expert resource for Glycogenomics', *Nucleic Acids Res* 37 (2008), D233-238.

17. S Kadowaki, K Yamamoto, M Fujisaki, K Izumi, T Tochikura, and T Yokoyama, 'Purification and characterization of a novel fungal endo-beta-N-acetylglucosaminidase acting on complex oligosaccharides of glycoproteins', *Agricultural and Biological Chemistry*, 54 (1990), 97-106.

18. T Kato, K Fujita, M Takeuchi, K Kobayashi, S Natsuka, K Ikura, H Kumagai, and K Yamamoto, 'Identification of an endo-beta-N-acetylglucosaminidase gene in Caenorhabditis elegans and its expression in Escherichia coli', *Glycobiology*, 12 (2002), 581-587.

19. K Takegawa, K Yamabe, K Fujita, M Tabuchi, M Mita, H Izu, et al., 'Cloning, sequencing, and expression of Arthrobacter protophormiae endo-beta-N-acetylglucosaminidase in Escherichia coli', *Archives of Biochemistry and Biophysics*, 338 (1997), 22-28

20. D W Abbott, M S Macauley, D J Vocadlo, and A B Boraston, 'Streptococcus pneumoniae endohexosaminidase D, structural and mechanistic insight into substrate-assisted catalysis in family 85 glycoside hydrolases', *The Journal of Biological Chemistry*, 284 (2009), 11676-11689

21. A L Tarentino, and F Maley, 'Purification and properties of an endo-beta-N-acetylglucosaminidase from Streptomyces griseus', *The Journal of Biological Chemistry*, 249 (1974), 811-817.

22. R B Trimble, and A L Tarentino, 'Identification of distinct endoglycosidase (endo) activities in Flavobacterium meningosepticum: endo F1, endo F2, and endo F3. Endo F1 and endo H hydrolyze only high mannose and hybrid glycans', *The Journal of Biological Chemistry*, 266 (1991), 1646-1651.

23. M Collin, and A Olsén, 'EndoS, a novel secreted protein from Streptococcus pyogenes with endoglycosidase activity on human IgG', *The EMBO Journal*, 20 (2001), 3046-3055

24. I Stals, B Samyn, K Sergeant, T White, K Hoorelbeke, A Coorevits, B Devreese, M Claeyssens, and K Piens, 'Identification of a gene coding for a deglycosylating enzyme in Hypocrea jecorina', *FEMS Microbiology Letters*, 303 (2010), 9-17.

25. T Hamaguchi, T Ito, Y Inoue, T Limpaseni, P Pongsawasdi, and K Ito, 'Purification, characterization and molecular cloning of a novel endo-beta-N-acetylglucosaminidase from the basidiomycete, Flammulina velutipes', *Glycobiology*, 20 (2010), 420-432.

26. V A Reddy, R S Johnson, K Biemann, R S Williams, F D Ziegler, R B Trimble, and F Maley, 'Characterization of the glycosylation sites in yeast external invertase. I. N-linked oligosaccharide content of the individual sequons', *The*

Journal of Biological Chemistry, 263 (1988), 6978-6985.

27. R B Trimble, and P H Atkinson, 'Structural heterogeneity in the Man8-13GlcNAc oligosaccharides from log-phase Saccharomyces yeast: a one- and two-dimensional 1H NMR spectroscopic study', *Glycobiology*, 2 (1992), 57-75.

28. M C Hertzberg, 'Purification, crystallization and structure determination studies of the enzymes Cel5A, Cel5B and EndoT from *Hypocrea jecorina*' Bachelor thesis advanced level E, *Swedish University of Agricultural Sciences (SLU)*, (2008).

29. K Fujita, R Nakatake, K Yamabe, A Watanabe, Y Asada, and K Takegawa, 'Identification of amino acid residues essential for the substrate specificity of Flavobacterium sp. endo-beta-N-acetylglucosaminidase', *Bioscience, Biotechnology, and Biochemistry*, 65 (2001), 1542-1548.

30. C A Waddling, T H Plummer, A L Tarentino, and P Van Roey, 'Structural basis for the substrate specificity of endo-beta-N-acetylglucosaminidase F(3)', *Biochemistry*, 39 (2000), 7878-7885.

31. T Bergfors, Editor. 'Protein Crystallization: Techniques, Strategies, and Tips', *International University Line, La Jolla, California* (1999), 300 pp.

32. J Drenth, 'Principles of protein X-ray crystallography', *Springer-Verlag, New York* (1994), 290 pp.

33. Dionex, 'Technical note TN20: analysis of carbohydrates by anion exchange chromatography with pulsed amperometric detection', Dionex, Sunnyvale (1989), USA.

34. N H Packer, M A Lawson, D R Jardine, and J W Redmond, 'A general approach to desalting oligosaccharides released from glycoproteins', *Glycoconjugate Journal*, 15 (1998), 737-747.

35. M Penttilä, H Nevalainen, M Rättö, E Salminen, and J Knowles, 'A versatile transformation system for the cellulolytic filamentous fungus Trichoderma reesei', *Gene*, 61 (1987), 155-164.

36. J Zhang, 'Structure-function studies of biomass degrading enzymes' Bachelor thesis advanced level E, *Swedish University of Agricultural Sciences (SLU)*, (2009). <http://stud.epsilon.slu.se/593/>

37. K Klarskov, K Piens, J Ståhlberg, P B Høj, J V Beeumen, and M Claeyssens, 'Cellobiohydrolase I from Trichoderma reesei: identification of an active-site nucleophile and additional information on sequence including the glycosylation pattern of the core protein', *Carbohydrate Research*, 304 (1997), 143-154.

38. B Bower, K Kodama, B Swanson, T Fowler, H Meerman, K Collier, C Mitchinson, and M Ward, 'Hyperexpression and glycosylation of *Trichoderma reesei* EG III', In: M Claeyssens, W Nerinckx, and K Piens, (Eds.), 'Carbohydrases from *Trichoderma reesei* and other microorganisms: structures, biochemistry, genetics and applications', *Royal Society of Chemistry, Cambridge* (1998), 327–334.

39. H Nevalainen, M Harrison, D Jardine, N Zachara, M Paloheimo, P Suominen, A Gooley and N Packer, 'Glycosylation of cellobiohydrolase I from *Trichoderma reesei*', In: M Claeyssens, W Nerinckx, and K Piens, (Eds.), 'Carbohydrases from *Trichoderma reesei* and other microorganisms: structures, biochemistry, genetics and applications', *Royal Society of Chemistry, Cambridge* (1998), 335–344.

40. J P M Hui, P Lanthier, T C White, S G McHugh, M Yaguchi, R Roy and P Thibault, 'Characterization of cellobiohydrolase I (Cel7A) glycoforms from extracts of Trichoderma reesei using capillary isoelectric focusing and electrospray mass spectrometry', *Journal of Chromatography B: Biomedical Sciences and Applications*, 752 (2001), 349-368.

41. I Stals, K Sandra, S Geysens, R Contreras, J V Beeumen, and M Claeyssens, 'Factors influencing glycosylation of Trichoderma reesei cellulases. I: Postsecretorial changes of the O- and N-glycosylation pattern of Cel7A', *Glycobiology*, 14 (2004), 713-724.

42. Rao, Vibha, Chudi Guan, and Patrick Van Roey, 'Crystal structure of endo-[beta]-N-acetylglucosaminidase H at 1.9 å resolution: active-site geometry and substrate recognition', *Structure*, 3 (1995), 449-457.

43. P V Roey, V Rao, T H Plummer, and A L Tarentino, 'Crystal Structure of Endo-.beta.-N-acetylglucosaminidase F1, an .alpha./.beta.-Barrel Enzyme Adapted for a Complex Substrate', *Biochemistry*, 33 (1994), 13989-13996.

44. A J McCoy, R W Grosse-Kunstleve, P D Adams, M D Winn, L C Storoni and R J Read, 'Phaser crystallographic software', *J. Appl. Cryst.*, 40 (2007), 658-674.

45. R B Trimble, A L Tarentino, T H Plummer, and F Maley, 'Asparaginyl

glycopeptides with a low mannose content are hydrolyzed by endo-beta-N-acetylglucosaminidase H', *The Journal of Biological Chemistry*, 253 (1978), 4508-4511.

46. A G W Leslie, 'Recent changes to the MOSFLM package for processing film and image plate data', *Joint CCP4 + ESF-EAMCB Newsletter on Protein Crystallography*, 26 (1992).

47. Collaborative Computational Project Number 4, 'The CCP4 suite: programs for protein crystallography', *Acta Crystallog. sect. D* 50 (1994), 760–763.

48. P Evans, 'Scaling and assessment of data quality', *Acta Crystallographica. Section D, Biological Crystallography*, 62 (2006), 72-82.

49. G N Murshudov, A A Vagin and E J Dodson, 'Refinement of macromolecular structures by the maximum-likelihood method', *Acta Crystallog. sect. D* 53 (1997), 240–255.

50. P Emsley, and K Cowtan, 'Coot: model-building tools for molecular graphics', *Acta Crystallographica. Section D, Biological Crystallography*, 60 (2004), 2126-2132.

51. W L DeLano, 'The PyMOL molecular graphics system', *Shrödinger LLC, New York, NY, USA* (2002). http://www.pymol.org

52. A C Terwisscha van Scheltinga, S Armand, K H Kalk, A Isogai, B Henrissat, och B W Dijkstra, 'Stereochemistry of chitin hydrolysis by a plant chitinase/lysozyme and X-ray structure of a complex with allosamidin: evidence for substrate assisted catalysis', *Biochemistry*, 34 (1995), 15619-15623.

53. M S Macauley, G E Whitworth, A W Debowski, D Chin, D J Vocadlo, '*O*-GlcNAcase uses substrate-assisted catalysis: kinetic analysis and development of highly selective mechanism-inspired inhibitors', *J Biol Chem.* 280 (2005), 25313–22.

54. S Knapp, D J Vocadlo, Z Gao, B Kirk, J Lou and S G Withers, 'NAG-thiazoline, an *N*-acetyl-β-hexosaminidase inhibitor that implicates acetamido participation', *J Am Chem Soc.* 118 (1996), 6804–6805.

55. S Drouillard, S Armand, G J Davies, C E Vorgias and B Henrissat, '*Serratia marcescens* chitobiase is a retaining glycosidase utilizing substrate acetamido group participation', *Biochemical Journal.* 328 (1997), 945–949.

56. S Arming, B Strobl, C Wechselberger and G Kreil, '*In vitro* mutagenesis of PH-20 hyaluronidase from human sperm', *Eur J Biochem.* 247 (1997), 810–814.